

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:12:11 ; Search time 30.32 Seconds  
(without alignments)  
1161.293 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVSCAPALGGCLGFTFS.....AQQGALRAPSQSGAAARS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	317	20 AAY06482	Human tumour-assoc
2	1720	100.0	317	20 AAY13391	Amino acid sequenc
3	1720	100.0	317	21 AAB11700	Human serine prote
4	1720	100.0	317	21 AAY93689	Amino acid sequenc
5	1720	100.0	317	22 AAY72890	Human serine prote
6	1720	100.0	317	22 AAB80259	Human PRO343 prote
7	1720	100.0	325	22 AAU23215	Novel human enzyme
8	1720	100.0	325	22 AAU17037	Human novel secret
9	1720	100.0	351	22 AAU16966	Human novel secret
10	1714	99.7	322	22 AAU23751	Novel human enzyme
11	1714	99.7	322	22 AAU17038	Human novel secret

12	1706	99.2	319	21 AAB11701	Human serine prote
13	1626	94.5	306	21 AAB11702	Human serine prote
14	1484	86.3	271	19 AAW77302	Amino acid sequenc
15	1444.5	84.0	327	22 AAY72891	PFEX-C-E-HIS fusio
16	1310	76.2	302	21 AAB11708	Human serine prote
17	1157.5	67.3	308	21 AAB11709	Mouse serine prote
18	1074	62.4	234	21 AAB11706	Human serine prote
19	1047	60.9	207	21 AAB11704	Human serine prote
20	1017	59.1	231	22 AAG75586	Human serine prote
21	739	43.0	290	21 AAY73388	HTRM clone 3376404
22	739	43.0	290	22 AAM39388	Human polypeptide
23	739	43.0	290	22 AAU12282	Human PRO4327 poly
24	739	43.0	290	22 AAE03821	Human gene 4 encod
25	739	43.0	290	22 AAB73945	Human protease T.
26	734	42.7	328	22 AAM41174	Human polypeptide
27	730	42.4	297	19 AAW77304	Amino acid sequenc
28	680.5	39.6	315	22 AAB73946	Fusion protein of
29	645	37.5	131	21 AAB11705	Human serine prote
30	628.5	36.5	284	22 AAU00467	Human serine prote
31	585.5	34.0	316	22 AAU00468	Recombinant human
32	582.5	33.9	319	21 AAB36481	Fusion gene with h
33	582.5	33.9	319	22 AAB67541	Amino acid sequenc
34	581	33.8	328	21 AAB36480	Fusion gene with h
35	581	33.8	328	22 AAB67540	Amino acid sequenc
36	554	32.2	276	19 AAW64242	Murine mast cell p
37	554	32.2	276	19 AAW63172	Mouse mast cell pr
38	549	31.9	658	22 AAE06934	Human membrane-typ
39	549	31.9	802	22 AAE06933	Human membrane-typ
40	547	31.8	802	20 AAY41710	Human PRO618 prote
41	547	31.8	802	21 AAB42466	Human PRO618 (UN03
42	547	31.8	802	21 AAB24052	Human PRO618 prote
43	538.5	31.3	274	19 AAW64234	Rat homologue of m
44	536.5	31.2	314	19 AAW77297	Amino acid sequenc
45	534	31.0	235	22 AAE06932	Human membrane-typ

## ALIGNMENTS

RESULT 1	
AAV06482	
ID	AAV06482 standard; Protein; 317 AA.
XX	
AC	AAV06482;
XX	
DT	27-SEP-1999 (first entry)
XX	
DE	Human tumour-associated protein PRO343.
XX	
KW	PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human.
XX	
OS	Homo sapiens.
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..32 "signal peptide"
FT	Protein 33..317
FT	/note= "mature protein"
FT	Modified-site 70
FT	/note= "N-glycosylated"
XX	
PN	WO9935170-A2.
XX	
PD	15-JUL-1999.
XX	
PF	05-JAN-1999; 99WO-US00106.
XX	
PR	20-NOV-1998; 98US-0109304.
PR	05-JAN-1998; 98US-0070440.
PR	29-APR-1998; 98US-0083500.
PR	22-MAY-1998; 98US-0086414.
PR	10-JUN-1998; 98US-0088742.
PR	10-NOV-1998; 98US-0107783.

XX (GETH ) GENENTECH INC.  
 PA Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;  
 PI Roy MA, Wood WI;  
 XX  
 XX WPI, 1999-430385/36.  
 DR N-PSDB; AAX87259.  
 XX  
 PT Antibody against proteins expressed in neoplastic cells, useful for  
 PT tumor diagnosis and treatment  
 XX  
 XX Example 1; Fig 12; 162pp; English.  
 XX  
 CC This sequence represents human PRO343 (UNQ302), a protein encoded  
 CC by the novel cDNA clone DNA43318 (see AAX87259). Amplification of  
 CC DNA43318 was observed in primary lung tumours and in primary colon  
 CC tumours, suggesting a significant role in tumour formation and  
 CC growth. Antagonists (e.g. antibodies) directed to PRO343 may have  
 CC use in cancer therapy. The invention identifies 14 genes (see  
 CC AAX87254-67) that are amplified in the genome of tumour cells. Such  
 CC amplification is expected to be associated with overexpression of  
 CC the gene product and to contribute to tumorigenesis. The encoded  
 CC proteins (see AAX06477-90) may be useful targets for the diagnosis  
 CC and/or treatment (including prevention) of certain cancers, and may  
 CC act as predictors of the prognosis of tumour treatment. Antibodies  
 CC that bind the proteins are claimed and used in claimed cancer  
 CC diagnostic kits.  
 XX  
 SQ Sequence 317 AA;  
 Query Match 100.0%; Score 1720; DB 20; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MVVSGAPPALGGCGTFTSLLLASTAILNAARIPVPPACGKPOQLNRVVGGEDSTDSE 60  
 DB 1 mvvsgappalggcgctftslilastailnaaripvpacgkppqqlnrvvvggedstdse 60  
 QY 61 WPVIVSIQKNGTHRCAGSLTSTRVITAACHCFKDNLNKPYLFLSGAWOLGNPGRSQK 120  
 DB 61 wpvivsIQKngthrcagslltstrvitaachcfkdnlnkpylflsGawqlngnpgsrsk 120  
 QY 121 VGVAWVEHPVYSKEGACADIALVRLESTQFSERVLPICLPDASIHLPNTHCWSGW 180  
 DB 121 vgvawvehpvyskEGacadiAlvrleSTqfservlPiclPdasiHlpnthcwsGw 180  
 QY 181 GSIDGVPPLPHTLQKLPIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACLG 240  
 DB 181 gsIdgvPlphtlQkLpIdseVcsHlywrgAggptedmLcagylegerdAcLg 240  
 QY 241 DSGGPLMCOVDGAWLLAGIISWGSGCAERNRPGVYISLSAHRSWKEIVQGVQLRGAQG 300  
 DB 241 dsggplmcvDgawllAGiISwGsgcaErnRpgvYiSlsAhrswkEivqGvqlrGaQg 300  
 QY 301 GGALRAPSQSGGAARS 317  
 DB 301 ggalrapsgsggaars 317  
 RESULT 2  
 AAY13391  
 ID AAY13391 standard; Protein; 317 AA.  
 XX  
 XX AAY13391;  
 AC  
 XX  
 DT 25-JUN-1999 (first entry)  
 XX  
 DE Amino acid sequence of protein PRO343.  
 XX  
 KW Secreted protein; transmembrane protein; human; enterocolitis;  
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;

KW congenital microvillus atrophy; skin disease; cell growth;  
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
 KW anti-thrombotic; wound healing; tissue repair.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9914328-A2.  
 XX  
 PD 25-MAR-1999.  
 XX  
 XX 16-SEP-1998; 98WO-US19330.  
 XX  
 XX 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059113.  
 PR 17-SEP-1997; 97US-0059113.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 17-SEP-1997; 97US-0059119.  
 PR 17-SEP-1997; 97US-0059121.  
 PR 17-SEP-1997; 97US-0059122.  
 PR 17-SEP-1997; 97US-0059184.  
 PR 18-SEP-1997; 97US-0059283.  
 PR 18-SEP-1997; 97US-0059286.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 21-OCT-1997; 97US-0063486.  
 PR 24-OCT-1997; 97US-0062814.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 24-OCT-1997; 97US-0063045.  
 PR 24-OCT-1997; 97US-0063120.  
 PR 24-OCT-1997; 97US-0063121.  
 PR 24-OCT-1997; 97US-0063127.  
 PR 27-OCT-1997; 97US-0063128.  
 PR 27-OCT-1997; 97US-0063329.  
 PR 28-OCT-1997; 97US-0063327.  
 PR 28-OCT-1997; 97US-0063541.  
 PR 28-OCT-1997; 97US-0063542.  
 PR 28-OCT-1997; 97US-0063544.  
 PR 28-OCT-1997; 97US-0063549.  
 PR 28-OCT-1997; 97US-0063550.  
 PR 28-OCT-1997; 97US-0063564.  
 PR 29-OCT-1997; 97US-0063435.  
 PR 29-OCT-1997; 97US-0063704.  
 PR 29-OCT-1997; 97US-0063732.  
 PR 29-OCT-1997; 97US-0063738.  
 PR 29-OCT-1997; 97US-0063734.  
 PR 29-OCT-1997; 97US-0064215.  
 PR 31-OCT-1997; 97US-0063735.  
 PR 31-OCT-1997; 97US-0063870.  
 PR 31-OCT-1997; 97US-0064103.  
 PR 03-NOV-1997; 97US-0064248.  
 PR 07-NOV-1997; 97US-0064809.  
 PR 12-NOV-1997; 97US-0065186.  
 PR 17-NOV-1997; 97US-0065846.  
 PR 18-NOV-1997; 97US-0065693.  
 PR 21-NOV-1997; 97US-0066120.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 24-NOV-1997; 97US-0066772.  
 PR 24-NOV-1997; 97US-0066466.  
 PR 24-NOV-1997; 97US-0066770.  
 PR 24-NOV-1997; 97US-0066511.  
 PR 24-NOV-1997; 97US-0066453.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
 PI WPI, 1999-229533/19.  
 DR N-PSDB; AAX52262.  
 DR  
 XX New isolated human genes and polypeptides used in, e.g. treatment of



```

RESULT      4
AA93689
ID  AA93689 standard; Protein; 317 AA.
XX
XX
AC  AA93689;
XX
XX  03-OCT-2000 (first entry)
XX
XX  Amino acid sequence of novel polypeptide PRO343.
XX
XX  PRO201; PRO292; PRO327; PRO1265; PRO344; PRO343; PRO347; PRO357;
XX  PRO715; PRO1017; PRO1112; PRO509; PRO853; PRO882; tumour cell;
XX  tumorigenesis; cancer; neoplastic cell growth; cell proliferation.
XX
XX  Homo sapiens.
XX
XX  Key      Location/Qualifiers
XX  Peptide  1..32
XX  Region   /note= "signal sequence"
XX  Modified-site 5..11 /note= "prokaryotic membrane lipoprotein attachment site"
XX  Modified-site 12..18 /note= "N-myristoylation site"
XX  Modified-site 13..19 /note= "N-myristoylation site"
XX  Modified-site 16..22 /note= "N-myristoylation site"
XX  Modified-site 52..58 /note= "N-myristoylation site"
XX  Modified-site 70..74 /note= "N-myristoylation site"
XX  Modified-site 71..77 /note= "N-glycosylation site"
XX  Modified-site 77..83 /note= "N-myristoylation site"
XX  Modified-site 86..92 /note= "N-myristoylation site"
XX  Active-site /note= "active site for serine protease"
XX  Modified-site 112..118 /note= "N-myristoylation site"
XX  Region     178..182 /note= "glycosylation attachment site"
XX  Modified-site 273..279 /note= "N-myristoylation site"
XX  Modified-site 310..316 /note= "N-myristoylation site"
XX
XX  WO200037640-A2.
XX
XX  29-JUN-2000.
XX
XX  16-DEC-1999; 99WO-US30095.
XX
XX  22-DEC-1998; 98US-0113296.
XX  08-MAR-1999; 99WO-US05028.
XX  02-JUN-1999; 99WO-US12252.
XX  01-SEP-1999; 99WO-US20111.
XX  15-SEP-1999; 99WO-US21090.
XX  30-NOV-1999; 99WO-US28313.
XX  01-DEC-1999; 99WO-US28409.
XX  02-DEC-1999; 99WO-US28301.
XX
XX  (GETH ) GENENTECH INC.
XX
XX  Botstein D, Goddard A, Gurney AL, Hillan K, Lawrence DA, Roy MA;
XX  Wood WI;
XX
XX  WPI: 2000-452188/39.
XX  N-PSDB; AAA46914.

```

```

XX  New anti-polypeptide antibody useful in the treatment and diagnosis of
XX  neoplastic cell growth and proliferation -
XX
XX  Claim 61; Fig 12; 220pp; English.
XX
XX  The present sequence represents a novel human polypeptide. The
XX  specification describes novel polypeptides designated PRO201, PRO292,
XX  PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO715, PRO1017,
XX  PRO1112, PRO509, PRO853 and PRO882. These genes are amplified in
XX  the genome of tumour cells. The polypeptides are believed to contribute
XX  to tumorigenesis. The polypeptides are useful target for the
XX  identification of certain cancers, and may act as predictors of the
XX  prognosis of tumour treatment. Antibodies against these polypeptides
XX  are useful in the treatment and diagnosis of neoplastic cell growth
XX  and proliferation in mammals.
XX
XX  Sequence 317 AA;
XX
XX  Query Match 100.0%; Score 1720; DB 21; Length 317;
XX  Best Local Similarity 100.0%; Pred. No. 3.2e-147;
XX  Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 MVVSGAPPALGGGCLGTFTSLLLASTAILNAARIPVPPACGKPKQLNRVVGEDSTDSE 60
XX  DB 1 mvvsgappalgggclgtftslillastailnaaripvpacgkpgqlnrvggedstdse 60
XX
XX  QY 61 WPWIVSIQKNGTHHCAGSLTSTRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGSRSQK 120
XX  DB 61 wpwivsiqkngthhcagslltstrwvitaahcfkdnlnkpylfsvllgawqlgnpgsrsk 120
XX
XX  QY 121 VGVAWVPHVPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASTHLPTNTHCWISGW 180
XX  DB 121 vgvaawvphvpvyswkegacadiavrlersiqfservlpiclpdasthlptnthcwisgw 180
XX
XX  QY 181 GSTQDGVPLPHPQTLQKLKVPITIDSEVCSHLYWRGAGQGPITEDMLCAGYLEGERDACLG 240
XX  DB 181 gsiqdgvpplphpqtllqklkvpildsevcshlywrgagggpitedmlcagylegerdaclg 240
XX
XX  QY 241 DSGGPLMCOVDGAWLLAGIISWEGGCAERNRPGVYISLAHRSWVERKIVQGVQLRGAQG 300
XX  DB 241 dsqgplmcvqdgawllagiisweggcaernrpgvyislahrswwerkivqgvqlrgragq 300
XX
XX  QY 301 GGALRAPSGGGAARS 317
XX  DB 301 ggalrapsgsgaaars 317
XX
XX  RESULT 5
XX  AA972890
XX  ID  AA972890 standard; Protein; 317 AA.
XX
XX  AC  AA972890;
XX
XX  DT  31-MAY-2001 (first entry)
XX
XX  DE  Human serine protease, protease C-E.
XX
XX  KW  Human; serine protease; protease C-E; therapy; desquamation; skin care;
XX  laundry detergent; shampoo; cleaning agent; hair care; skin flaking;
XX  neurodegenerative disorder; dermatological; immunogenic; proteolytic.
XX  chromosome 16p13.3.
XX
XX  OS  Homo sapiens.
XX
XX  PN  WO200116288-A2.
XX
XX  PD  08-MAR-2001.
XX
XX  PF  14-AUG-2000; 2000WO-US22117.
XX
XX  PR  31-AUG-1999; 99US-0386529.

```

XX (ORTH ) ORTHO-MCNEIL PHARM INC.  
 XX Darrow A, Qi J, Andrade-Gordon P;  
 XX WPI; 2001-226681/23.  
 DR N-PSDB; AAD02990.  
 XX Novel serine protease termed protease C-E, useful for treating and  
 PT preventing skin flaking or imbalance of desquamation -  
 XX Claim 11; Fig 1; 78pp; English.  
 XX The present sequence is a human serine protease, protease C-E which  
 CC is a member of the SI serine protease family. Protease C-E gene is  
 CC located on chromosome 16p13.3 and is expressed in pancreas, placenta,  
 CC prostate, small intestine, stomach, spleen, fibroblasts, epidermis,  
 CC cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is  
 CC useful for treating an imbalance of desquamation, by topical application.  
 CC A non-pharmaceutical composition comprising the protein may be formulated  
 CC as a laundry detergent, shampoo, hard surface cleaning composition, dish  
 CC care cleaning composition, skin care composition and hair care  
 CC composition. Protease C-E is useful for treating and preventing skin  
 CC flaking, neurodegenerative disorders and dermatological pathologies. It  
 CC is less immunogenic to sensitive individuals and it provides efficient  
 CC proteolytic activity in a non-natural environment.  
 XX SQ Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSGAPPALGGGCLGTTFTSLLLASTAILNAARIPVPACGKPKQLNRVVGEDSTDSE 60  
 DB 1 mvsgappalgggclgtftstlllastailnaaripvpacgkpkqlnrvggedstdse 60

QY 61 WPWIVSTQKNGTHHCASLLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQLNGPGRSQK 120  
 DB 61 wpwivstqkngthhcaglltsrwwitaahcfkdnlmkpylfsvllgawqlngpgrsqk 120

QY 121 VGVAWPEHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASITLHPNTHCWISGW 180  
 DB 121 vgvaewpehpvyswkegacadiavrlersiqfservlpiclpdasihlpnthcwisgw 180

QY 181 GSIQDGVPLPHPOTLQKLKVPIDSEVCSHLYWRGAGQGPTEDMLCAGYLEGERDACL 240  
 DB 181 gsiqdgvpplphpptlqlkvpildsevcshlywrgaggptedmlcagylegerdaclg 240

QY 241 DSGGPLMCOVDGAWLLAGIISWGEGCAERNRPGVYISLSAHSRWKIVQGVQLRGAQG 300  
 DB 241 dsggplmcvqdgawllagilswgegcaernrpgvyislsahrsrwkivqgvqlrgragq 300

QY 301 GGALRAPSQSGGAARS 317  
 DB 301 ggalrapsgsggaars 317

RESULT 6  
 AAB80259  
 ID AAB80259 standard; Protein; 317 AA.  
 XX  
 AC AAB80259;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Human PRO343 protein.  
 XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;  
 KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiac;  
 KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;  
 KW antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;

opthalmological; gene therapy; skin disease; gastrointestinal disorder;  
 ischaemia; inflammation.  
 Homo sapiens.  
 WO200104311-A1.  
 18-JAN-2001.  
 22-FEB-2000; 2000WO-US04414.  
 07-JUL-1999; 99US-0143048.  
 26-JUL-1999; 99US-0145698.  
 28-JUL-1999; 99US-0146222.  
 08-SEP-1999; 99WO-US20594.  
 13-SEP-1999; 99WO-US20944.  
 15-SEP-1999; 99WO-US21090.  
 15-SEP-1999; 99WO-US21547.  
 03-OCT-1999; 99WO-US23089.  
 23-NOV-1999; 99WO-US28214.  
 30-NOV-1999; 99WO-US28313.  
 16-DEC-1999; 99WO-US30095.  
 20-DEC-1999; 99WO-US30911.  
 20-DEC-1999; 99WO-US30999.  
 03-JAN-2000; 99WO-US00219.  
 (GETH ) GENENTECH INC.  
 Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;  
 Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;  
 Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;  
 Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;  
 Williams PM, Wood WI;  
 WPI; 2001-081051/09.  
 N-PSDB; AAF72420.  
 Sixty one nucleic acids encoding PRO polypeptides which are useful in  
 the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung  
 squamous cell carcinoma) and neurodegenerative diseases (e.g.  
 Alzheimer's disease) -  
 Claim 1; Fig 98; 393pp; English.  
 The present sequence is one of sixty one novel secreted and  
 transmembrane PRO polypeptides. The PRO polypeptides are  
 useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
 squamous cell carcinoma), gastrointestinal disorders (e.g.  
 enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
 Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
 endometrial bleeding angiogenesis, ischaemias such as coronary  
 ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
 rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
 diabetes and retinal disorders such as retinitis pigmentosum.  
 The PRO nucleic acids have applications in molecular biology, including  
 use as hybridization probes, and in chromosome and gene mapping.

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSGAPPALGGGCLGTTFTSLLLASTAILNAARIPVPACGKPKQLNRVVGEDSTDSE 60  
 DB 1 mvsgappalgggclgtftstlllastailnaaripvpacgkpkqlnrvggedstdse 60

QY 61 WPWIVSTQKNGTHHCASLLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQLNGPGRSQK 120  
 DB 61 wpwivstqkngthhcaglltsrwwitaahcfkdnlmkpylfsvllgawqlngpgrsqk 120

QY 121 VGVAWPEHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASITLHPNTHCWISGW 180





PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229387.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256179.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 03-JAN-2001; 2001US-0254097.  
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476222/51.

N-PSDB; AAS26942.

Novel polypeptides and polynucleotides useful as diagnostic reagents to  
diagnose diseases or disorders associated with aberrant expression or  
activity of polypeptides, for treating blood clotting disorder,  
haemophilia

Claim 11; SEQ ID No 278; 601pp; English.

The invention relates to isolated nucleic acid molecules and their  
encoded secreted proteins. The nucleic acids and proteins are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They  
are also used in diagnosing a pathological condition or susceptibility  
to a pathological condition. Antibodies to the proteins can also  
be used in alleviating symptoms associated with the disorders and in  
diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
include autoimmune diseases e.g. rheumatoid arthritis,  
hyperproliferative disorders e.g. neoplasms of the breast or liver,  
cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
Alzheimer's disease, infections caused by bacteria, viruses and fungi  
and ocular disorders e.g. corneal infection, and many other



CC disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present	
Query Match 100.0%; Score 1720; DB 22; Length 325;	
Best Local Similarity 100.0%; Pred. No. 3.3e-147;	
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy 1	MVSGAPPALGGCGLTGFTSLLLASTAILNAARIPVPACGKPOQLNRVYGGEDSTDSE 60
Db	
Qy 9	mvsappalggcgltgftslslllastailnaaripvpacgkpoqlnrvyggedstdse 68
Db	
Qy 61	WPWIVSIQKNGTHHCAGSLLTSRWVITAACHFKDNLNKPFLFSVLLGAWQLGNPGRSQK 120
Db	
Qy 69	wpwivsiqkngthhcagslltsrwvitaahcfkdnlnkpylfsvllgawqlgnpgsrsqk 128
Db	
Qy 121	VCVAVVEHPVYSWKEGACADIALVRLRSIQFSERVLPICLPDASIHLPNTHCWISGW 180
Db	
Qy 129	vgvavvehpvyswkegacadiavrlrsiqfservlpiclpdasihlpnthcwisgw 188
Db	
Qy 181	GSIQGVPLPQTPQTKLKPITIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACLG 240
Db	
Qy 189	gsiqgvplpqtqtklkvpidevcsghlywrgaggptedmlcagylegerdaclg 248
Db	
Qy 241	DSGGPLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHSRWVKIVQGVQLRGAQQ 300
Db	
Qy 249	dsggplmcqvdgawllagliswegcaernrpgvyislsahrsrwkivqgvqlrgragg 308
Db	
Qy 301	GCALRAPSGSGAARS 317
Db	
Qy 309	ggalrapsqsgaars 325
Db	
RESULT 9	
AAU16966	
ID	AAU16966 standard; Protein; 351 AA.
AC	AAU16966;
XX	
DT	07-NOV-2001 (first entry)
DE	Human novel secreted protein, SEQ ID 207.
XX	
KW	Human; immunosuppressive; antiarthritic; antirheumatic;
KW	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW	vulnary; secreted protein; rheumatoid arthritis;
KW	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW	cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW	nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW	corneal infection; wound healing; epithelial cell proliferation;
KW	skin ageing; food additive; preservative; antiproliferative.
OS	Homo sapiens.
XX	
PN	WO200155441-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01320.
XX	
XX	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.

PR 19-MAY-2000;	2000US-0205515.
PR 07-JUN-2000;	2000US-0209467.
PR 28-JUN-2000;	2000US-0214886.
PR 30-JUN-2000;	2000US-0215135.
PR 07-JUL-2000;	2000US-0216647.
PR 07-JUL-2000;	2000US-0216880.
PR 11-JUL-2000;	2000US-0217487.
PR 11-JUL-2000;	2000US-0217496.
PR 14-JUL-2000;	2000US-0218290.
PR 26-JUL-2000;	2000US-0220963.
PR 26-JUL-2000;	2000US-0220964.
PR 14-AUG-2000;	2000US-0224518.
PR 14-AUG-2000;	2000US-0224519.
PR 14-AUG-2000;	2000US-0225213.
PR 14-AUG-2000;	2000US-0225214.
PR 14-AUG-2000;	2000US-0225266.
PR 14-AUG-2000;	2000US-0225267.
PR 14-AUG-2000;	2000US-0225268.
PR 14-AUG-2000;	2000US-0225270.
PR 14-AUG-2000;	2000US-0225447.
PR 14-AUG-2000;	2000US-0225757.
PR 14-AUG-2000;	2000US-0225758.
PR 14-AUG-2000;	2000US-0225759.
PR 18-AUG-2000;	2000US-0226279.
PR 22-AUG-2000;	2000US-0226681.
PR 22-AUG-2000;	2000US-0226686.
PR 22-AUG-2000;	2000US-0227182.
PR 30-AUG-2000;	2000US-0227009.
PR 01-SEP-2000;	2000US-0228924.
PR 01-SEP-2000;	2000US-0229287.
PR 01-SEP-2000;	2000US-0229343.
PR 01-SEP-2000;	2000US-0229344.
PR 01-SEP-2000;	2000US-0229345.
PR 05-SEP-2000;	2000US-0229509.
PR 05-SEP-2000;	2000US-0229513.
PR 06-SEP-2000;	2000US-0230437.
PR 08-SEP-2000;	2000US-0230438.
PR 08-SEP-2000;	2000US-0231242.
PR 08-SEP-2000;	2000US-0231243.
PR 08-SEP-2000;	2000US-0231244.
PR 08-SEP-2000;	2000US-0231413.
PR 08-SEP-2000;	2000US-0231414.
PR 08-SEP-2000;	2000US-0232080.
PR 08-SEP-2000;	2000US-0232081.
PR 12-SEP-2000;	2000US-0231968.
PR 14-SEP-2000;	2000US-0232397.
PR 14-SEP-2000;	2000US-0232398.
PR 14-SEP-2000;	2000US-0232399.
PR 14-SEP-2000;	2000US-0232400.
PR 14-SEP-2000;	2000US-0232401.
PR 14-SEP-2000;	2000US-0233063.
PR 14-SEP-2000;	2000US-0233064.
PR 14-SEP-2000;	2000US-0233065.
PR 21-SEP-2000;	2000US-0234223.
PR 25-SEP-2000;	2000US-0234274.
PR 25-SEP-2000;	2000US-0234997.
PR 26-SEP-2000;	2000US-0234998.
PR 26-SEP-2000;	2000US-0235484.
PR 27-SEP-2000;	2000US-0235834.
PR 29-SEP-2000;	2000US-0235836.
PR 29-SEP-2000;	2000US-0236327.
PR 29-SEP-2000;	2000US-0236367.
PR 29-SEP-2000;	2000US-0236368.
PR 29-SEP-2000;	2000US-0236369.
PR 29-SEP-2000;	2000US-0236370.
PR 02-OCT-2000;	2000US-0236802.
PR 02-OCT-2000;	2000US-0237037.
PR 02-OCT-2000;	2000US-0237038.
PR 02-OCT-2000;	2000US-0237039.
PR 13-OCT-2000;	2000US-0237040.
PR 13-OCT-2000;	2000US-0239935.
PR 20-OCT-2000;	2000US-0239937.
PR 20-OCT-2000;	2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PR XX  
 FA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-476222/51.  
 DR N-PSDB; AAS26871.  
 DR  
 XX Novel polypeptides and polynucleotides useful as diagnostic reagents to  
 PT diagnose diseases or disorders associated with aberrant expression or  
 PT activity of polypeptides, for treating blood clotting disorder,  
 PT haemophilia  
 XX  
 XX Claim 11; SEQ ID NO 207; 60lpp; English.  
 PS  
 XX The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC  
 Query Match 100.0%; Score 1720; DB 22; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-147; Indels 0; Gaps 0;  
 Matches 317; Conservative 0; Mismatches 0;  
 Qy 1 MVSGAPPALGGGCGTFTSLLLASTAILNAARIPVPACGKPKQLNRVVGSDTDSE 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 35 mvsgappalgggclgtftslillastailnaaripvpacgkpkqlnrvggedstdse 94  
 Qy 61 WPWIVSTQKNGTHHCAGSLTSTRWVITAAHCFKONLKNPYLFLSVLLGAWQLNGPGRSQK 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 95 wpwivsiqkngthhcagslltsrwvitaahcfkdnlnkpylflsvllgawqlnpgsrsk 154  
 Qy 121 VGVAVVEPHVPVYSKNEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWSGW 180  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 155 vgvavvephvpvyswkegacadiavrlersiqfservlpiclpdasihlpnthcwsigw 214  
 Qy 181 GSIQDGVPLPHPQTLQKLVPIIDSEVCSHLYWRGAGGPITEDMLCAGYLEGRDACLG 240  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 215 gsiqdgvpplphpqtlqlkvpildsevcshlywrgaggpitedmlcagylegrdacig 274  
 Qy 241 DSGGPLMCQVDGANLLAGIISWEGGCAERNRPGVYISLSAHSRWEKIVQGVQLRGAQG 300  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 275 dsggplmcqvdgawllagilswgegcaernrpgvyislsahrswekivqgvqlrgragq 334  
 Qy 301 GGALRAPSQSGAAARS 317  
 Db ||||||||||||||||  
 335 ggalrapsgsgaaars 351  
 RESULT 10  
 AAU23751  
 ID AAU23751 standard; Protein; 322 AA.  
 XX  
 XX AAU23751;  
 AC  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 XX Novel human enzyme polypeptide #837.  
 DE  
 XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
 KW autoimmune disorder; neurological disorder; metabolic disorder;  
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
 KW nephrotropic; anticoagulant.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200155301-A2.  
 PN  
 XX  
 XX 02-AUG-2001.  
 PD





PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-476222/51.  
XX N-PSDB; AAS26943.  
XX  
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
PT diagnose diseases or disorders associated with aberrant expression or  
PT activity of polypeptides, for treating blood clotting disorder,  
PT haemophilia  
XX  
XX Claim 11; SEQ ID No 279; 601pp; English.  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present

Query Match 99.7%; Score 1714; DB 22; Length 322;  
Best Local Similarity 99.7%; Pred. No. 1.1e-146;  
Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGGCLGTTFTSLLLLASTAILNAARIPVPACGKPPQQLNRVVGGEDSTDSE 60  
Db 6 mvvsgappaigggclgtftstlllastailnaaripvpacgkppqqlnrvvvggedstdse 65  
QY 61 WPWIVSIQKNGTHHCAGSLTTSRWVITAAHCFKDNLNKPYLFSVLLGAWOLNPGRSRQK 120  
Db 66 wpwivsixkngthhcagsltstwttaahcfkdnlnkpylfsvllgawolnpgrsrsk 125  
QY 121 VGVAVVEPHVPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCHWISGW 180  
Db 126 vgvavvephvpvyswkegacadiavrlersiqfservlpiclpdasihlpnthchwisgw 185  
QY 181 GSTODGVPLPHTLOKLYPIIDSEVCSHLYWFGAGGPTEDMLCAGYLEGERDACLIG 240  
Db 186 gsiqdgvpplphptlqklypiidsevcshlywrgagggpitedmlcagylegerdaci 245  
QY 241 DSGGLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVGQVLRGAAQ 300  
Db 246 dsggplmcqvdbgawllagiiswegcaernrpgvylsahrswekivgqvlgrraqq 305  
QY 301 GGALRAPSGSGAARS 317  
Db 306 ggalrapsqsgsaars 322



```
CC murine BSSP4 (mBSSP4). Sequences AAB11700-B11708 represent human BSSP4
CC variants (hBSSP4), and sequence AAB11709 represents murine BSSP4
XX
SQ Sequence 306 AA;

Query Match 94.5%; Score 1626; DB 21; Length 306;
Best Local Similarity 97.4%; Pred. No. 9.5e-139;
Matches 300; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 1 MVVSGAPPALGGGCLGFTTSLLLASTAILNAARIPVPACGKPKQQLNRVVGSDTDSE 60
DB 1 mvvsgappalgggclgfttslllascailnaaripvpacgkpkqqlnrvvgsdtdse 60
QY 61 WPIVTSIQKNGTHRCAGSLTSTRVITAACHCKDNLNKPYLFSVLLGAWQLGNFGSRSQK 120
DB 61 wpivtsiqkngthrcagslltsrvtitaachckdnlnkpylfsvllgawqlgnfgsrsqk 120
QY 121 VGVAVPEHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASTHLPNTHCWSGW 180
DB 121 vgvavpehpvyswkegacacadiavrlersiqfservlpiclpdasthlpnthcwsgw 180
QY 181 GSIQGVPLPHPQTLQKLKVPIDSEVCSHLYWRGAGOGPITEDMLCAGYLEGERDACL 240
DB 181 gsidgvplphpqtlqklkvpidsescshlywrgagggpitedmlcagylegerdac 240
QY 241 DSGGGLMCQVDGAWLLAGIISWGEGCAERNRPYVYISLSAHSRVERIVQGVQLRGAQG 300
DB 241 dsggplmcqvdgawllagiiSWGEGCAERNRPYVYISLSAHSRVERIVQGVQLRGA 300
QY 301 GGALRAP 308
DB 299 ----rapa 302

RESULT 14
AAW77302
ID AAW77302 standard; Protein: 271 AA.
AC AAW77302;
XX
DT 07-JAN-1999 (first entry)
XX
DE Amino acid sequence of SP0011A, a homologue of HELA2.
XX
KW Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KW BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;
KW seminoma; testis-specific expression; antitumour; sperm development;
KW infertility; human; chromosome 16p13.3.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 29
FT /note= "likely to be involved in disulphide bonding"
FT Misc-difference 44
FT /note= "likely to be a catalytic residue"
FT Disulfide-bond 45
FT /note= "likely to be involved in disulphide bonding"
FT Misc-difference 95
FT /note= "likely to be a catalytic residue"
FT Disulfide-bond 115
FT /note= "likely to be involved in disulphide bonding"
FT Disulfide-bond 129
FT /note= "likely to be involved in disulphide bonding"
FT Misc-difference 160
FT /note= "encoded by GGAA"
FT Disulfide-bond 162
FT /note= "likely to be involved in disulphide bonding"
FT Disulfide-bond 181
FT /note= "likely to be involved in disulphide bonding"
FT Disulfide-bond 192
```

```
FT Misc-difference 196
FT /note= "likely to be involved in disulphide bonding"
FT Disulfide-bond 202
FT /note= "likely to be a catalytic residue"
FT Disulfide-bond 220
FT /note= "likely to be involved in disulphide bonding"
FT Misc-difference 235
FT /note= "likely to be involved in disulphide bonding"
XX WO9836054-A1.
XX 20-AUG-1998.
XX 13-FEB-1998; 98WO-AU00085.
XX 18-NOV-1997; 97AU-0000422.
XX 13-FEB-1997; 97AU-0005101.
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX Antalis TM, Hooper JD;
XX WPI: 1998-480768/41.
XX N-PSDB; RAV59134.
XX
XX New serine protease(s) and kinase involved in regulating cell
XX activity and viability - particularly the testis-specific protease
XX HELA2 used for modulation of fertility and as tumour suppressor
XX
XX Example 15; Fig 20A; 167pp; English.
XX
XX AAW77302-04 represent HELA2 homologues. The genes are found in a cluster
XX on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
XX homology to serine proteases. The protein is involved in or associated
XX with regulation of cell activity and/or viability. Administration of
XX recombinant HELA2 (also called testisin) is used to increase fertility.
XX Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
XX testicular germ cell cancers (seminoma) and is also expressed in some
XX non-testicular cancers (of colon, pancreas, prostate and ovary), so is
XX a marker/potential therapeutic target for cancer. The promoter from the
XX HELA2 gene is useful for testis-specific expression of other genes,
XX e.g. for gene therapy or modulation of fertility. Drugs that block
XX activity of HELA2 should have antitumour activity (other than in
XX testis) while in testis recombinant HELA2 should stop growth of tumours
XX and normalise sperm development (eliminating the need for orchidectomy).
XX Identification of mutant forms of HELA2 can be used to diagnose
XX infertility.
XX
XX Sequence 271 AA;
```

```
Query Match 86.3%; Score 1484; DB 19; Length 271;
Best Local Similarity 100.0%; Pred. No. 5.5e-126;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 LNRVVGSDTDSEMPWIVTSIQKNGTHRCAGSLTSTRVITAACHCKDNLNKPYLFSVLL 106
DB 1 lnrvvgsdtdsempwivtsiqkngthrcagslltsrvtitaachckdnlnkpylfsvll 60
QY 107 GAWQLGNFGSRSQKVGAVPEHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDAS 166
DB 61 gawqlgnfgsrsqkvavpehpvyswkegacacadiavrlersiqfservlpiclpdas 120
QY 167 IHLPPNTHCWSGWISQGVPLPHPQTLQKLKVPIDSEVCSHLYWRGAGOGPITEDML 226
DB 121 ihlppnthcwsghisqgvplphpqtlqklkvpidsescshlywrgagggpitedml 180
QY 227 CAGYLEGERDACLDSGGGLMCQVDGAWLLAGIISWGEGCAERNRPYVYISLSAHSRWE 286
DB 181 cagylegerdacldsggplmcqvdgawllagiiSWGEGCAERNRPYVYISLSAHSRWE 240
QY 287 KIVGVQLRGAQGGALRAPSGSGAARS 317
```

Db 241 kivqgvqlrgragggalrapsqgsaaars 271

RESULT 15

AAV72891  
ID AAV72891 standard; Protein; 327 AA.

AA  
AC  
AA72891;AA  
AC  
AA72891;

DT 31-MAY-2001 (first entry)

DT 31-MAY-2001 (first entry)

AA PFEK-C-E-HIS fusion protein.

AA PFEK-C-E-HIS fusion protein.

Human; serine protease; protease C-E; therapy; desquamation; skin care;  
 KW  
 laundry detergent; shampoo; cleaning agent; hair care; skin flaking;  
 KW  
 neurodegenerative disorder; dermatological; immunogenic; proteolytic;  
 KW  
 bovine; zymogen activation construct; PFEX2-C-E-HIS ERI-HCII;  
 KW  
 fusion protein; chromosome 16p13.3.  
 KW

Chimeric - Bos sp.

OS Chimeric - Homo sapiens.

XX  
PN WO200116288-A2.

08-MAR-2001.

14-AUG-2000; 2000WO-US22117.

31-AUG-1999; 99US-0386629.

PA (ORTH ) ORTHO-MCNEIL PHARM INC.

PI Darrow A, Qi J, Andrade-Gordon P;

WPI; 2001-226681/23.

DR N-PSDB; AAD02991.

Novel serine protease termed protease C-E, useful for treating and preventing skin flaking or imbalance of desquamation -

PS Claim 11; Fig 4; 78pp; English.

The present sequence is PREK-C-E-HIS fusion protein encoded by a zymogen activation construct, PREK2-C-E-HIS ERI-HCII. The construct comprises bovine preproelastin signal sequence fused in-frame with MoAbM2 anti-FLAG antibody epitope for the purpose of secretion and antibody detection (PF), enterokinase cleavage site from human trypsinogen I (EK), catalytic domain of protease C-E and six histidine codons (6XHIS). Protease C-E gene located on chromosome 16p13.3 is a member of the S1 serine protease family and is expressed in pancreas, placenta, prostate, small intestine, stomach, spleen, fibroblasts, epidermis, cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is useful for treating an imbalance of desquamation, by topical application. A non-pharmaceutical composition comprising the protein may be formulated as a laundry detergent, shampoo, hard surface cleaning composition, dish care cleaning composition, skin care composition and hair care composition. Protease C-E is useful for treating and preventing skin flaking, neurodegenerative disorders and dermatological pathologies. It is less immunogenic to sensitive individuals and it provides efficient proteolytic activity in a non-natural environment.

Sequence 327 AA;

Query Match	84.0%;	Score 1444.5;	DB 22;	Length 327;
Best Local Similarity	87.1%;	Pred. No. 2.6e-122;		
Matches 271;	Conservative	7;	Mismatches 14;	Indels 19; Gaps 2;

21 LLLLASTAIL-----NAARIPVPACGKPPQQLNRVVGGEDSTDSEWPWIVS 66

```

14 llllvsnlllcqgvsvdykdddddvdlaalaap-----fdddkivggvaldedsewpwiwv 68
Db

```

Qy.	67	IQKNGTHHCAGSLLTSRWVTTAAHCFKDNLNKPYLFSVILGAWQLGNPQSRSQKVGAVW	126
Db	69	iqkngthcagslltsrwwttaahcfkdnlnkpylfsvilgawqlgnpqsrsqkvavw	128
Qy	127	EPHPVYSWKEGACADIALVLERSIQFSERVLPICLPDASIHLPNTHCWSIGSWGSIQDG	186
Db	129	ephpvyswkegacadiatlvrslsqfservlpiclpdasihlpnthcwsigswsiqdg	188
Qy	187	VPHPQTLQKLKVPDIIDSEVCSHLYWRGAGQGPTIETDMLCAGYLSGERDACLGDSGGPL	246
Db	189	vpiphpqtllqklkvpdiidsevcshlywrgagqgptiedmlcagylegerdaciIdsggpl	248
Qy	247	MCQVDGAWLLAGIISWEGECAERNRPGVYIISLSAHSWSWEKIVQGVQLRGAQGGGALRA	306
Db	249	mcqvdbgawllagiiswgegaernrpgvyiislsahrswekivqgvqlrgaqgggallra	308
Qy	307	PSQSGSGAAARS	317
Db	309	psqsgsgaaars	319

Search completed: August 29, 2002, 10:24:08  
Job time: 717 sec



THIS PAGE BLANK (USPTO)

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:12:26 ; Search time 12.99 Seconds  
(without alignments)  
596.067 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVVSGAPALGGCLGTFTS.....AOGGALRAPSGCGAAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581.5	33.8	299	4	US-08-944-483-66
2	554	32.2	276	2	US-09-016-366A-15
3	554	32.2	276	2	US-08-978-404B-21
4	538.5	31.3	274	2	US-08-978-404B-5
5	531.5	30.9	314	4	US-09-008-271A-3
6	530.5	30.8	274	2	US-09-016-366A-21
7	530.5	30.8	274	2	US-08-978-404B-16
8	530.5	30.8	275	2	US-09-016-366A-17
9	530.5	30.8	275	2	US-08-978-404B-12
10	529.5	30.8	273	2	US-09-016-366A-19
11	529.5	30.8	273	2	US-08-978-404B-14
12	522	30.3	270	2	US-08-978-404B-8
13	519.5	30.2	273	2	US-08-978-404B-6
14	518	30.1	273	2	US-08-978-404B-3
15	514.5	29.9	267	2	US-09-016-366A-23
16	514.5	29.9	267	2	US-08-978-404B-18
17	511.5	29.7	245	4	US-08-944-483-69
18	511.5	29.7	249	4	US-09-079-970A-5
19	508	29.5	304	4	US-09-088-651-2
20	506.5	29.4	245	4	US-09-079-970A-6
21	473.5	27.5	248	4	US-08-944-483-63
22	470.5	27.4	238	4	US-08-944-483-64
23	469.5	27.3	638	2	US-08-681-151-3
24	468.5	27.2	435	4	US-09-008-271A-6
25	466	27.1	454	4	US-09-518-046-2
26	463	26.9	416	2	US-09-000-846-2
27	462	26.9	492	4	US-09-342-749-2

28	453	26.3	283	3	US-08-807-151-1	Sequence 1, Appli
29	453	26.3	283	4	US-09-478-957-1	Sequence 1, Appli
30	449	26.1	791	1	US-08-643-219-1	Sequence 1, Appli
31	449	26.1	791	3	US-08-851-350-1	Sequence 1, Appli
32	447	26.0	814	1	US-08-750-711-1	Sequence 1, Appli
33	446	25.9	546	6	5200340-6	Patent No. 5200340
34	446	25.9	790	1	US-08-469-486-54	Sequence 54, Appl
35	446	25.9	790	2	US-08-469-658-54	Sequence 54, Appl
36	446	25.9	791	2	US-09-131-995-1	Sequence 1, Appli
37	446	25.9	791	2	US-08-832-087B-1	Sequence 1, Appli
38	446	25.9	791	4	US-09-132-154-1	Sequence 1, Appli
39	446	25.9	810	1	US-07-854-603-2	Sequence 2, Appli
40	446	25.9	810	1	US-08-147-000B-29	Sequence 29, Appl
41	446	25.9	810	4	US-09-086-514-1	Sequence 1, Appli
42	446	25.9	810	6	5200340-8	Patent No. 5200340
43	445.5	25.9	256	2	US-09-027-337-3	Sequence 3, Appli
44	443.5	25.8	798	1	US-08-200-900A-2	Sequence 2, Appli
45	443.5	25.8	798	5	PCT-US94-00616-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-944-483-66  
; Sequence 66, Application US/08944483  
; Patent No. 6232456  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STEWART, KENT D.  
; APPLICANT: STROUPE, STEVEN D.  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,483  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6183.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: No. 6232456e  
US-08-944-483-66

Query Match 33.8%; Score 581.5; DB 4; Length 299;  
Best Local Similarity 45.4%; Pred. No. 6.5e-50;  
Matches 113; Conservative 42; Mismatches 89; Indels 5; Gaps 4;  
QY 50 VVGEDSTDEWPIVSIQKNGTHHCAGSLTTSRWITAAHCFKDNKPYLFSVLLGAW 109  
DB 1 ITGSSAVAGOWPQVSIYEGVHVCGLVSEQWVLSAAHCFSEHKE-AYEVKLGAAH 59  
QY 110 QLNPGRSRQKVGVAWPEHPVPSWKEGACADIALVRLERSIQFSERVLPICLPDASIH 169  
DB 60 QLDSEDAKVSTLKDIIHPSY-LQSGSGDIALQLSRPITPSRYIRPICLPAANASE 118  
QY 170 PPNTHCWISGWSIQDGVPLPHQTLQKLVPIIDSEVCSHLYWRGA-GOGP--ITEDML 226  
DB 119 PNGLHCTVTGWHVAPSVSLTPKPLQOLEVPLISRETCLYNIDAKPEEPHFVQEDMV 178  
QY 227 CAGYLEGERDACLDGSGGLMCOVDGAWLLAGIISWEGGCAERNRPGVWISLSAHSRWSVE 286  
DB 179 CAGIVEGGKDACQDGGSLPCPVEGLWYLTGIYSWGDACGARNRPGVYTLASSYASW 238  
QY 287 KIVGVOLR 295  
DB 239 SKVTELOPR 247

RESULT 2  
US-09-016-366A-15  
; Sequence 15, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; NUMBER OF INVENTIONS: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-016-366A-15

Query Match 32.2%; Score 554; DB 2; Length 276;  
Best Local Similarity 41.6%; Pred. No. 3.2e-47;  
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;  
QY 21 LLLIATAILNAARIPVP-PACGKPOOLNRVVGEDSTDEWPIVSIQ---KNGTHHCA 76  
DB 6 LLLWALSLSLASLYSAPRA-----NORVGIVGGHEASESKWPQVSLRFLKLTWIFCG 61  
QY 77 GSLTTSRWITAAHCFKDNKPYLFSVLLGAWOL--GNPGRSQKVGVAWPEHPVYSW 134  
DB 62 GSLTHPQWLTAACHVGHPIKSPOLFRVQLREQYLYYGD-----QLLSLNRIVVHPHYT 116  
QY 135 KEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWSIQDGVPLPHQ 194  
DB 117 AEGG-ADVALLELEVNVNTHIHPISLPASETFFPGTSCWVTGWGDIDNDEPLPPYP 175  
QY 195 LQKLKVPIDSEVCSHLYWRGAGOG---PITED-MLCAGYLEGERDACLDGSGGLMCOV 250  
DB 176 LKQVKVPIVENSLCDRKYHTGLYTGDDFPVHDGMLCAG--NTRDSCQDGGSLVCKV 233  
QY 251 DGAWLLAGIISWEGGCAERNRPGVWISLSAHSRWSVEKIV 289  
DB 234 KGTWLQAGVSWGEGCAQPNKPGIYTRVYLDWIHRV 272

RESULT 3  
US-08-978-404B-21  
; Sequence 21, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e  
; US-08-978-404B-21

Query Match 32.28; Score 554; DB 2; Length 276;  
Best Local Similarity 41.68; Pred. No. 3.2e-47;  
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;

Qy 21 LLLASTAAILNAARIPVP-PACGRPQQLNRVVGGEDSTDSEWPIVSIQNGT---KNGTHCA 76  
Db 6 LLLWALSILASLYSAPRPA---NORVIGVGHESKWPQVSLURFKLNTWIHFPG 61

Qy 77 GSLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQL--GNPGSRSQKYGVAWVEPHPVYSW 134  
Db 62 GSLHPQWLTAACHVGHPIKSPQLFRVQLREQVLYYGD-----QLLSLNRIVVPHYYT 116

Qy 135 KEGACADIALYRLERSIQFSERVLPICLPDASIHLPNTHCWISGSIQDGVLPHPQT 194  
Db 117 AEGG-ADVALLELEVNVNTHIPLSPASETFPGTSCWVTGWGIDNDEPLPPYP 175

Qy 195 LQKLKVPIDSEVCSHLYWRGAGOG---PITD-MLCAGYLEGERDACLGSDGGPLMCOV 250  
Db 176 LKQVKVPIVENSICDRKYHTGLYTGDDFPVHDGMLCAG--NTRDSCQDGGGFLVCKV 233

Qy 251 DGAWLLAGISWGEGCAERNRPGVYISLSAHRSWVEKIV 289  
Db 234 KGTWLAGVWSWEGCAQPNKPGIYTRVYYLDWIHRYV 272

## RESULT 4

US-08-978-404B-5  
; Sequence 5, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 274 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e

Query Match 31.38; Score 538.5; DB 2; Length 274;

Best Local Similarity 40.68; Pred. No. 1.1e-45;  
Matches 117; Conservative 46; Mismatches 84; Indels 41; Gaps 10;

Qy 21 LLLAST---AILNAARIPVPACGKPKQQLNRVVGGEDSTDSEWPIVSIQNGT---HH 74  
Db 5 LLLALSPLASLVHAACPVKQVRG-----IVGGREASESKWPQVSLRKFESFWMHF 57

Qy 75 CAGSLTSRWVITAAHCFKDNLNKPYLFSVLLGAWQLGNPGSRSQ-----KVGVAW 125  
Db 58 CGSLIHPQWLTAACHVGLHIKSPQLFRVQL-----REQVLYYADQLLTVNRTV 107

Qy 126 VEPHPVYSWGEGACADIALYRLERSIQFSERVLPICLPDASIHLPNTHCWISGSIQD 185  
Db 108 VHPH-YITVEDG--ADIALLELEIPNVNTHIPLSPASETFPGTSCWVTGWGIDDS 164

Qy 186 GVPLPHPQTLQKLVPIIDSEVCSHLYWRGAGOG---PITD-MLCAGYLEGERDACLG 241  
Db 165 DEPLPPYPLQVKVPVIVENSICDRKYHTGLYTGDDVPVQDGMCLCAG--NTRDSCQGD 222

Qy 242 SGGPLMCOVDCAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIV 289  
Db 223 SGGPLVCKVKTWLAGVWSWEGCAEANREGIYTRVYYLDWIHRYV 270

## RESULT 5

US-09-008-271A-3  
; Sequence 3, Application US/09008271A  
; Patent No. 6203979  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Hillman, Jennifer L.  
; Yue, Henry  
; Guegler, Karl J.  
; Corley, Neil C.  
; Tang, Tom Y.  
; Shah, Purvi  
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/008,271A  
; FILING DATE: 16-Jan-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mohan-Peterson, Sheela  
; REGISTRATION NUMBER: 41,201  
; REFERENCE/DOCKET NUMBER: PF-0458 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 314 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSTUT03  
; CLONE: 789927  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-008-271A-3

Query Match 30.9%; Score 531.5; DB 4; Length 314;  
Best Local Similarity 36.0%; Pred. No. 6.5e-45;  
Matches 111; Conservative 57; Mismatches 115; Indels 25; Gaps 6;  
QY 13 GCLTFTSLLLASTAIL--NAARIPVPACGPPQQLNRVVGEDSTDSWPMIVSIQK 69  
Db 2 GARGALLALLARAGLRKPKESQAAPLSGPGRRVITSRIVGGEAELGRWPQGSRL 61  
QY 70 NTHHCAGSLTSRWVITAAHCRK--DNLNKPILFSLVLLGA-----WOLGNPGSRSQK 120  
Db 62 WDSRVCGVSLLSHRWALTAAHCFETYSDLSDPGMMVQFGOLTSMPFSWLQAYYTR-YF 120  
QY 121 VGVAWVPHVPVYSKGEACADIALVRLERSIQFSERVLPICLPDASIHLPENTHCWISGW 180  
Db 121 VSNILSPR---YLGNSPDYIALVKLSAPVYTKHQPICLQASTFEFNRTDCWGTGW 176  
QY 181 GSIODGVPPLPQTLQKLVPIIDSEVCSHLYWRGAGOGPITEDMLCAGYLEGERDACLG 240  
Db 177 GYIKEDALSPHTLQEVQVAIINSMCNHLFLKYSFRKDFGDMVCAGNAQGGKDACFG 236  
QY 241 DSGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQ 300  
Db 237 DSGPLACNKNGLMYQIGVYVSWGCGRPNRPGVYTNISHFEWIOKLM-----AQS 288  
QY 301 GGALRAP 308  
Db 289 GMSQDPFS 296

RESULT 6

US-09-016-366A-21  
; Sequence 21, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 274 amino acids  
; TYPE: amino acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-016-366A-21

Query Match 30.8%; Score 530.5; DB 2; Length 274;  
Best Local Similarity 39.2%; Pred. No. 6.8e-45;  
Matches 113; Conservative 53; Mismatches 85; Indels 37; Gaps 9;  
QY 20 SLLLASTAILNAARIPVPACGPPQQLNRVVGEDSTDSWPMIVSIQKNG---THHCA 76  
Db 2 NLLLLALPVL--ASRAYAAPGQALQRVGIVGGQAPRSKWPQVSLRVHGPYWMHFCG 59  
QY 77 GSLTSRWVITAAHCRK--FKD-----NLNKPILFSLVLLGAWQLGNPGSRSQKVGVAWV 126  
Db 60 GSLTHPQWLTAHCVGPDVKDLAALRVQLREQHLY-----YQQLLPVSRI 106  
QY 127 EPHP-VYSWEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGSIQD 185  
Db 107 IVHPQFYTAQIG--ADIALLELEBPVKVSSHVHTVTLPPASETEPPGMPGCVWTGMDVDN 164  
QY 186 GVPLPHPOTLQKLVPIIDSEVCSHLYWRGAGOGP-----ITEDMLCAGYLEGERDACLGD 241  
Db 165 DERLPPFPPLQKQVPIIMENHICDAKYLHGAITGDDVRIVRDDMLCAG--NTRRDSQGD 222  
QY 242 SGGPLMCQVDGAWLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIV 289  
Db 223 SGGPLCVCKNGTWTLOAGVSVSGEGCAQPNRPGIYTRVYIYLDWIHHYV 270

RESULT 7

US-08-978-404B-16  
; Sequence 16, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; TITLE OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 274 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5968782e

9:50

	Matches	113;	Conservative	52;	Mismatches	86;	Indels	37;	Gaps	9;										
Qy	20	SLLLLA	STAILNAARI	PVPPACGR	PQQLNR	VVVGCGD	STDS	SENPWIS	IQKNG---	THHCA 76										
		: : : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :										
Db	1	NLLLL	ALLPVL--	ASRAYAAP	GGQALOR	GVGTG	QGEAPR	SKRKPWQ	VSRLRV	GHGPPYMHFFCG 58										
Qy	77	GSLT	SRWVITAAHC	---FKD---	NLNK	PYLEF	SVLLG	AWQLGN	PGSRQ	SRQGVAVW 126										
		: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :										
Db	59	GSLI	HPQWLVTAAHC	VGPDVKD	LAALR	VQLR	REQHLY	-----	YDQLLP	VSRI 105										
Qy	127	EPHP	--VYSW	KEGACAD	IATL	VALRS	IQFSE	RVLPIC	LPDASIH	LPNTHC	WISGWSIQD 185									
		: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :										
Db	106	IVH	QFYTAQIG-	-ADAT	ALLEE	PEPVN	SSHVTV	TLP	PA	SEFP	PPGMP	CVWTVGWGDVDN 163								
Qy	186	GVPL	PHPQT	LQKLK	VIID	SEVCS	HLV	WRGAGQ	GP----	ITE	DM	L	CAGYLE	GEDRAC	L	IGD 241				
		: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :				
Db	164	DERL	PPPP	PPKQ	VKVP	MPIN	ENHIC	DAK	YHLG	AYTGDD	YRIV	VRD	DM	L	CAG--	N	TRRDS	CQGD 221		
Qy	242	SGGL	PLMCQ	VDG	AW	L	LAGI	ISW	GEC	CAERN	PGV	YI	S	I	S	AHRS	SW	VEKTV 289		
		: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :	: : :		
Db	222	SGGL	PLV	KV	NGT	VLQ	AGV	SVS	WG	CAQ	PNR	PGI	Y	TR	Y	T	Y	Y	LDW	THHYV 269

RESULT 11  
 US-08-978-404B-14  
 : Sequence 14, Application US/08978404B  
 : Patent No. 5968782  
 : GENERAL INFORMATION:  
 : APPLICANT: Stevens, Richard L.  
 : TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 : TITLE OF INVENTION: FIBRINOGEN  
 : NUMBER OF SEQUENCES: 74  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 : STREET: 600 Atlantic Avenue  
 : CITY: Boston  
 : STATE: MA  
 : COUNTRY: U.S.A.  
 : ZIP: 02210-2211  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: DOS  
 : SOFTWARE: FastSeq for Windows Version 2.0  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/978,404B  
 : FILING DATE: 25-NOV-97  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 60/032,354  
 : FILING DATE: 04-DEC-1996  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Plumer, Elizabeth R.  
 : REGISTRATION NUMBER: 36,637  
 : REFERENCE/DOCKET NUMBER: B0801/7090  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 617-720-3500  
 : TELEFAX: 617-720-2441  
 : TELEX:  
 : INFORMATION FOR SEQ ID NO: 14:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 273 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: No. 5968782e  
 US-08-978-404B-14

Query Match 30.8%; Score 529.5; DB 2; Length 273;  
Best Local Similarity 39.2%; Pred. No. 8.5e-45;  
Matches 113; Conservative 52; Mismatches 86; Indels 37; Gaps 9;





Qy	78	SLITSRWVITAAHCXFDN-----LNKPYLFSVLIGANQLGNPGRSOKVGVAWVE	127
		::   ::     ::   ::   ::	
Db	60	SLHPQWLVITAAHCVGPKADPNKLVRQLRKQYLY-----YHDLLTAVSQIIL	106
Qy	128	PHP-VYSWKEGACADTALRLERSIOFSERVLPICLPDASIHLPPTHCHWISQWSIQDG	186
Db	107	SHPDFYIAQGQ--ADATALLKLTNPVNITSNVHTVSLPPASETFPSCTLCWCVTGWGMIND	164
Qy	187	VPLPHPTLOKLVPPIIDSVCSHLRYRGAQGP----ITEMLCAGYLEGERDACLGDS	242
Db	165	VSLEPPPLEEVQPIVENRLCDLKHYHGLTGNTGDNVHIVRDDMLCACN-EG-HDCQGD	222
Qy	243	GGPLMCQVDGANLAGLIISWGECGAERNRPVVYSISAHRSRWVEKITV	289
Db	223	GGPLVCVKEDTVLQAGVSWGECGAOPNRPGIYTRVTYLDWTIRYL	269

```

RESULT 14
US-08-978-404B-3
; Sequence 3, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIRINGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5968782e
US-08-978-404B-3

```

```

Db      61 LIHPQWVLTAAHCVGPDPVADPNKVRVLRKQYLY-----YHDHLMTVSQIIT 10
Qy      129 HP-VYSWKEGACADIALVRLERSIQFSERYLPICLPDASIHLPNTHCWISGWGSIQDGV 187
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      108 HPDFYIVQDG--ADIALLKLTNPVNSIDYVHPVPLPPASETFPSGTLCTWWTGNGNDGV 165
Qy      188 PLPHPTQLKLVPIIDSEVCSHLHYWEGACQP-----ITEDMLCAGYLSEGDACILGDSG 243
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      166 NLPPFPFLKEVQVPVPIENHLCCLDKYHKGLTIGDNVHIVRDMMLCAGN-EG-HDSCQGDG 223
Qy      244 GPLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSSWEKIV 289
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      224 GPLVCKVEDTQLAGVYVSWEGCAQPNRPGIYTRVYIYLDWIIHYV 269
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-016-366A-23
; Sequence 23, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
;   APPLICANT: Stevens, Richard L.
;   APPLICANT: Huang, Chifu
;   TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
;   TITLE OF INVENTION: INHIBITORS
;   NUMBER OF SEQUENCES: 65
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
;     STREET: 600 Atlantic Avenue
;     CITY: Boston
;     STATE: MA
;     COUNTRY: U.S.A.
;     ZIP: 02210-2211
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSeq for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/016,366A
;     FILING DATE: January 30, 1998
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/037,090
;     FILING DATE: 05-FEB-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Plumer, Elizabeth R.
;     REGISTRATION NUMBER: 36,637
;   REFERENCE/DOCKET NUMBER: B0801/7093
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-720-3500
;     TELEFAX: 617-720-2441
;     TELEX:
;   INFORMATION FOR SEQ ID NO: 23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 267 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;   US-09-016-366A-23

```

Qy 138 ACADIALVRLERSIQSERVLPICLPDASHLPENTHCWISGWSIODGYPLPHPTLQK 197  
Db 112 --ADIALLELEEPVKVSSHVTTLPPASETFPPGMPCWVTGWDVNDNDRLPPLKQ 169  
Qy 198 LKVPIDSEVCSHLYWRGAGQGP----ITEDMLCAGYLEGERDACLGDGSGGPLMCQVDGA 253  
Db 170 VKVPIMENHICDAKYHLGAYTGDDVRIVRDDMLCAG--NTRRDSQGDGSGGPLVCKVNGT 227  
Qy 254 WLLAGIISWEGGCAERNRPGVYISLSAHRSWVEKIV 289  
Db 228 WLQAGVYSWEGGCAQPNRPGIYTRVYIYLDWIHHYV 263

Search completed: August 29, 2002, 10:24:28  
Job time: 722 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:13:56 ; Search time 17.52 Seconds  
(without alignments)  
1738.603 Million cell updates/sec

Title: US-09-903-925a-263  
Perfect score: 1720  
Sequence: 1 MVVSGAPPALGGGCLGTFTS.....AQQGGALRAPSGSGAAARS 317  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	607.5	35.3	343	1 A57014	proctasin (EC 3.4.21.1)
2	554	32.2	276	2 A38654	mast cell proteina
3	546	31.7	275	2 A32410	tryptase (EC 3.4.21.1)
4	538.5	31.3	274	2 JC4171	tryptase (EC 3.4.21.1)
5	536	31.2	237	2 S68702	tryptase (EC 3.4.21.1)
6	532	30.9	274	2 A45754	tryptase (EC 3.4.21.1)
7	530.5	30.8	275	2 B35863	tryptase (EC 3.4.21.1)
8	529.5	30.8	275	2 A35863	tryptase (EC 3.4.21.1)
9	522.5	30.4	275	2 C35863	tryptase (EC 3.4.21.1)
10	522	30.3	270	2 S56160	mast cell tryptase
11	518	30.1	273	2 A47246	tryptase (EC 3.4.21.1)
12	481	28.0	638	1 KQHUP	plasma kallikrein
13	475.5	27.6	638	1 KFHUI	coagulation factor
14	473	27.5	625	1 KQMPL	plasma kallikrein
15	469.5	27.3	638	1 KQTPPL	plasma kallikrein
16	460.5	26.8	812	1 PLBO	plasmin (EC 3.4.21.1)
17	457.5	26.6	1034	1 A53663	enteropeptidase (E
18	452	26.3	417	1 S00845	hepsin (EC 3.4.21.1)
19	450	26.2	416	1 S33777	hepsin (EC 3.4.21.1)
20	450	26.2	810	2 B30848	plasmin (EC 3.4.21.1)
21	446	25.9	455	2 A61545	plasmin (EC 3.4.21.1)
22	446	25.9	790	1 PLPG	plasmin (EC 3.4.21.1)
23	446	25.9	810	1 PLHU	plasmin (EC 3.4.21.1)
24	445.5	25.9	1524	2 T30337	polypeptide - Afri
25	445	25.9	786	1 A47347	serine proteinase
26	444	25.8	1019	1 A56318	enteropeptidase (E
27	443.5	25.8	1035	1 A43090	enteropeptidase (E
28	442	25.7	1420	2 A32869	apolipoprotein(a)
29	441	25.6	460	2 B61545	plasmin (EC 3.4.21.1)

30	439.5	25.6	812	1 PLMS	plasmin (EC 3.4.21.1)
31	437.5	25.4	367	2 JE0104	testicular serine
32	437.5	25.4	437	2 S18407	acrosin (EC 3.4.21.1)
33	437.5	25.4	4548	1 S00657	apoptotrypsin(a) (EC 3.4.21.1)
34	437	25.4	263	1 KQRTB	chymotrypsin (EC 3.4.21.1)
35	437	25.4	421	1 S11674	acrosin (EC 3.4.21.1)
36	434	25.2	265	2 T15451	hypothetical prote
37	433.5	25.2	271	2 A25528	pancreatic elastas
38	432.5	25.1	436	2 JX0172	acrosin (EC 3.4.21.1)
39	430.5	25.0	366	2 JE0105	testicular serine
40	429.5	25.0	431	2 S47538	acrosin (EC 3.4.21.1)
41	428	24.9	263	2 A21195	chymotrypsin (EC 3.4.21.1)
42	425	24.7	810	2 I46260	plasmin (EC 3.4.21.1)
43	424.5	24.7	270	2 B29934	pancreatic elastas
44	422	24.5	263	2 A31299	chymotrypsin (EC 3.4.21.1)
45	422	24.5	415	1 A34170	acrosin (EC 3.4.21.1)

ALIGNMENTS

RESULT 1  
A57014  
proctasin (EC 3.4.21.1) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 18-Jun-1999  
C:Accession: A57014; A54866  
R;Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 270, 13483-13489, 1995  
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of  
A:Reference number: A57014; MUID:95286644  
A:Accession: A57014  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-343 <RES>  
A:Cross-references: GB:I41351; NID:g862304; PIDN:AAC41759.1; PID:g862305  
A:Experimental source: prostate  
A>Note: parts of this sequence were determined by protein sequencing  
R;Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 269, 18843-18848, 1994  
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purificatio  
A:Reference number: A54866; MUID:94308140  
A:Accession: A54866  
A:Molecule type: protein  
A:Residues: 45-64 <YUA>  
C:Genetics:  
A:Gene: GDB:PRSS8  
A:Cross-references: GDB:676446; OMIM:600823  
A:Map position: 16p11.2-16p11.2  
C:Superfamily: proctasin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-44,45-343/Product: proctasin #status predicted <MAT>  
F:33-44/Domain: proctasin light chain #status predicted <CHL>  
F:45-343/Domain: proctasin heavy chain #status predicted <CHH>  
F:45-281/Domain: trypsin homology <TRY>  
F:323-341/Domain: transmembrane #status predicted <TM1>  
F:37-154,70-86,168-244,201-223,234-262/Disulfide bonds: #status predicted  
F:85,134,238/Active site: His, Asp, Ser #status predicted  
F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match	35.3%	Score	607.5	DB	1	Length	343
Best Local Similarity	44.2%	Pred. No.	6.3e-44				
Matches	129	Conservative	42	Mismatches	108	Indels	13
						Gaps	7
QY	10	LGSGCLGTFSTLL---	LLASTAILNAARIPVPPACGKQQLNRVVGSGEDSTDSEMPWIVS	66			
Db	7	LGPGQLGAVALLYLGLLRSGTGAGAP---	CGVAPQA-RITGSSAVAGQMPWQVS	61			
QY	67	IQKNGTHHCAGSLTSRWVITAAHCFKDNLNKPYLFVLLGAWQLGNPGRSQKVGVAWV	126				
Db	62	ITYEGVHVGGSLSYSEQWLSAAHCFPSEHKE-AYEVKLGHAHQLDYSYSEDAKYSTLADI	120				

Qy	127	E P H P V T S W K E G A C A D I A L V R L E S I Q F S E R V I P U I C L P D S A I H L P P N T H C W I S G W S I O D G	186
		:     :         :         :     :     :	
Db	121	I P H P S Y - L Q E G S Q D I A L L Q L S R P I T F S R V I R P I C L P A A N A S F P N C L H T C T V T G W H V A P S	179
		:     :         :         :     :     :	
Qy	187	V P I P H P O T L Q K L V P I D S E V C H L Y W R G A - Q G P P -- I T E D M L C A G Y L E G E R D A C L G D S G	243
		:     :     :     :     :     :     :	
Db	180	V S L L T P K P L Q O E V P L I S R E T C N C L Y N I D A K P E P H F V Q E D M V C A G Y E G G D A C Q G D S G	239
		:     :     :     :     :     :     :     :	
Qy	244	G P L M C Q V D G A W L L A G I I S W G E C A E R N P C V Y I S L S A H R S W E K I V Q G V Q L R	295
		:     :     :     :     :     :     :     :	
Db	240	G P L S C P E G L W Y U T G I V S N G D A C A G R N R E G V Y T L A S S Y A S W I Q S K V T E L O P R	291
		:     :     :     :     :     :     :     :	

RESULT 2

A38654  
mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse  
C:Date: Mus musculus (house mouse)  
C:Update: 21-Feb-1992 #sequence\_revision 17-Feb-1994 #text\_change 22-Jun-1999  
C:Accession: A38654; B38654; D35646; I59478  
R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.; Serafin, W.E.  
J. Biol. Chem. 266, 3847-3853, 1991  
A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Trans  
A:Reference number: A38654; MUID:91139682  
A:Accession: A38654  
A:Molecule type: DNA  
A:Residues: 1-276 <REY>  
A:Cross-references: GB:IM57625; NID:g200506; PIDN:AAA39987.1; PID:g200507  
A:Note: the authors translated the codon CGC for residue 24 as Ala, GAG for  
s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for  
A:Accession: B38654  
A:Molecule type: mRNA  
A:Residues: 1-276 <REY>  
A:Cross-references: GB:IM57626; NID:g200508; PIDN:AAA39988.1; PID:g200509  
R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Sera  
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990  
A:Title: Different mouse mast cell populations express various combinations  
A:Reference number: A35646; MUID:90222202  
A:Accession: D35646  
A:Molecule type: protein  
A:Residues: 32-54 <RE3>  
R:Huang, R.; Abrink, M.; Gobi, A.E.; Nilsson, G.; Aveskogh, M.; Larsson, L.  
Scand. J. Immunol. 38, 359-367, 1993  
A:Title: Expression of a mast cell tryptase in the human monocytic cell line  
A:Reference number: I59478; MUID:94023807  
A:Accession: I59478  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-276 <RES>  
A:Cross-references: GB:L31853; NID:g473480; PIDN:AAA39725.1; PID:g473481  
C:Genetics:  
A:Gene: MMCP-6  
A:Introns: 24/1; 79/2; 168/1; 222/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-31/Domain: activation peptide #status predicted <ACT>  
F:32-276/Product: mast cell proteinase 6 #status experimental <MAT>  
F:32-268/Domain: trypsin homology <TRY>  
F:75,122,225/Active site: His, Asp, Ser #status predicted

Query Match	32.2%	Score 554;	DB 2;	Length 276;
Best Local Similarity	41.6%;	Pred. No. 1.7e-39;		
Matches 116;	Conservative 48;	Mismatches 93;	Indels 22;	Gaps
QY	21	LLLLASTATLNAARIPVP--PACGKFPQQLNRVVGGEDSTDSEWPNTVTSIQ----	KNGTTHCA	76
		: : :	: : :	:
Db	6	LLLLWALSLLSLVTSAPRPA----NQRVGIVGGHEASESKWPQVSLRFLNVTWIFCG	61	
QY	77	GSILTSRWVITAAHCFKONLKNPKPYLFSLVLLGANQL--CNPGRSQKVCVAVVEHPHYYSW	134	
		: : :	:	:
Db	62	GLSIHPQWVLTAAHCGVPHIKSPQFLRVQLREQVLYYGD----QLLSLNRIVVVPHYYT	116	

Qy	135	KEGACADIALVRLERSIQFSERVLPCLPDASIHLPNTHCWSIGWGSITQDGVPPLHPQOT	194
Dd	117	AEGG-ADVALLEEVPNVNSTHPIPSLPPASTFPFGTSCWVTGWGDINDDEPILPPYP	175
Qy	195	LQKLVIIDSEVCSHLYRWGAQG---PITED-MLCAGYLEGERDACILSGGGPLMCOV	250
Dd	176	LKQVKVIPENSLCDRKYHTGLTGDDFPIVDHMLCAG--NTRDRSCOGDSGGPLVCKY	233
Qy	251	DGAWLLAGIISWGEGAERNRPQVYISLSAHRSWVEKIY	289
Dd	234	KGTWLQAGVYSWGEGAOPNKPIQTVRYTLDWIRHYV	272

RESULT 3  
A32410  
trypsinase (EC 3.4.21.59) precursor - dog  
C:Species: Canis lupus familiaris (dog)  
C:Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 22-Jun-1999  
C:Accession: A32410  
R:Vanderslice, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.  
Biochemistry 28, 4148-4155, 1989  
A:Title: Molecular cloning of dog mast cell trypsinase and a related protease  
A:Reference number: A32410; MUID: 89352460  
A:Accession: A32410  
A:Molecule type: mRNA  
A:Residues: 1-275 <VAN>  
A:Cross-references: GB:M24664; NID:q163982; PIDN:AAA30854.1; PID:g163983; C:  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <STG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsinase #status predicted <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match	31.7%;	Score 546;	DB 2;	Length 275;
Best Local Similarity	40.9%;	Pred. No. 8e-39;		
Matches 112; Conservative	46;	Mismatches 102;	Indels 14;	Gaps
QY	23	LLASTALNAARIPVPACGKQQQLNRVVGGSDSDTSEMPWIVSIOKNG--TTHCAGSL	79	
Db	5	LVLALALLGSL-VPVSPAFQALQRVGIVGGREAPGSKWPQVSLRLKQGYWRHICGGS	63	
QY	80	LTSRWVITRAHCFKDLNLPKPYLSVLLGAWOLGNGPSRSQKVGVAWVPHPVYSHKEAC	139	
Db	64	IFPQWVLTTRAHCVPNGVNVCPPEEIRVOLREQHL---YYQDHLPLPNVRIVVHNPYYTPENG-	119	
QY	140	ADIALVRLERSTQFSRVLPICLPDASIHLPENTHCWISGSDIGDVPPLPHQPTLOKLK	199	
Db	120	ADIALLEEDPNVNSAHVQPVTLPPALQFTPTGTFCWVTGMDVHSGTPLPPPPPLKQVK	179	
QY	200	VPIIDSEVCSHLYWRGAGGPP---ITEDMLCAGYLEGERDACLGSDGSLMCQVDGANL	255	
Db	180	VPVIVEMSCDVQIHLGLSTGDGVRIREDMLCAG--NSKSDSCQDSDGSGPLVCRVRGVM	237	
QY	256	LAGIISWEGGCAERNRPVGIYISLSAHRSMVKIV	289	
Db	238	QAGVSWSEGCAOPNRPGIYTRVAYVLDWIHOV	271	

```

RESULT# 4
JC4171
  tryptase (EC 3.4.21.59) precursor - rat
N;Alternate names: mast cell tryptase
C:Species: rattus norvegicus (Norway rat)
C:Date: 27-Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000
C:Accession: JC4171
R:Ide, H.: Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.;
J. Biochem. 118, 210-215, 1995
A:Title: cDNA sequencing and expression of rat mast cell tryptase.
A:Reference number: JC4171; MUID:96015171

```



```

RESULT 7
A35863
tryptase (EC 3.4.21.59) II precursor - human
N:Alternate names: tryptase beta
C:Species: Homo sapiens (man)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: B35863; A37193; I59473
R:Vandersllice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr
A:Reference number: A35863; MUID:90251647
A:Accession: B35863
A:Molecule type: mRNA; DNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M33492; NID:g339982; PIDN:AAA36779.1; PID:g339983
A:Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic
R:Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A:Title: Cloning and characterization of a second complementary DNA for human tryptase.
A:Reference number: A37193; MUID:90369005
A:Accession: A37193
A:Molecule type: mRNA
A:Residues: 1-275 <MIL>
A:Cross-references: GB:M37488; NID:g179583; PIDN:AAA51843.1; PID:g179584
R:Blom, T.; Hellman, L.
Scand. J. Immunol. 37, 203-208, 1993
A:Title: Characterization of a tryptase mRNA expressed in the human basophil cell line K
A:Reference number: I59473; MUID:93166209
A:Accession: I59473
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-275 <RES>
A:Cross-references: GB:S55551; NID:g265666; PIDN:AAD13876.1; PID:g4261576
A:Experimental source: basophil cell line KU812
C:Genetics:
A:Gene: GDB:TPS1
A:Cross-references: GDB:125890; OMIM:191080
A:Map position: 16pter-16qter
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: trypsin I #status predicted <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.8%; Score 530.5; DB 2; Length 275;
Best Local Similarity 39.2%; Pred. No. 1.6e-37;
Matches 113; Conservative 53; Mismatches 85; Indels 37; Gaps 9;

QY 20 SLLLLASTAILNAARIPVPPACGKPGQQLNRVVGGEEDSTDSEWPWIVSIQKNG---THHCA 76
:|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 NLLLLALPVL--ASRAYAAPGALQGVGGQAPRKPQVSLRVHGPVWMHFCG 60
:|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 GSLLTSRWVITAHC-----NLNKPYLFVSVLLGAWQLGNPGRSRQKGVAVW 126
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 GSLLHPQWLTAACHVGPDKVLAALRVQLREQHLY-----YQDQLLPVSRI 107
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 127 EPHP-VYSWEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCHWISGWSIQD 185
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 IVHPQFYTAQIG--ADIALLEEFVNVSSHVHTVTLPPASSETFPFGMPGCVTWGSDVN 165
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 186 GVPLPHPTQKLKVPITDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLDG 241
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 DERLPPFPPLKQVQKVPIMENHICDAKYLHGYTGDDVRIYRDDMLCAG--NTRDSCQGD 223
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 SGGPLMCQVDGAWLLAGIISWEGEACERNRPGVYISLSAHRWSVEKTV 289
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 224 SGGPLVCKVNGTWLQAGVSWEGEACQPNRPGIYTRVTVYLDWIHHVY 271
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 8
A35863
tryptase (EC 3.4.21.59) I precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 21-Jul-2000
C:Accession: A35863; D35863; A60939; A39326
R:Vandersllice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey,
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine
A:Reference number: A35863; MUID:90251647
A:Accession: A35863
A:Molecule type: DNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC83172.1; PID:g339977
A:Accession: D35863
A:Molecule type: mRNA
A:Residues: 1-275 <VA2>
A:Cross-references: GB:M33491
R:Butterfield, J.H.; Weiler, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.
J. Leukoc. Biol. 47, 409-419, 1990
A:Title: Purification of tryptase from a human mast cell line.
A:Reference number: A60939; MUID:90244210
A:Accession: A60939
A:Molecule type: protein
A:Residues: 31-38, 'P', '40-41', 'X', '43', '45-48', 'X', '50 <BUT>
A:Experimental source: mast cell
A:Note: 44-Gly was also found
R:Cromlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretien
J. Biol. Chem. 262, 1363-1373, 1987
A:Title: Human pituitary tryptase: molecular forms, NH-2-terminal sequence, immunocy
A:Reference number: A39326; MUID:87109258
A:Accession: A39326
A:Molecule type: protein
A:Residues: 31-38 <CRO>
A:Experimental source: pituitary
C:Genetics:
A:Introns: 21/1; 78/2; 177/1; 221/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: trypsin I #status experimental <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.8%; Score 529.5; DB 2; Length 275;
Best Local Similarity 39.2%; Pred. No. 2e-37;
Matches 113; Conservative 52; Mismatches 86; Indels 37; Gaps 9;

QY 20 SLLLLASTAILNAARIPVPPACGKPGQQLNRVVGGEEDSTDSEWPWIVSIQKNG---THHCA 76
:|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 NLLLLALPVL--ASRAYAAPGALQGVGGQAPRKPQVSLRVHGPVWMHFCG 60
:|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 GSLLTSRWVITAHC-----NLNKPYLFVSVLLGAWQLGNPGRSRQKGVAVW 126
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 GSLLHPQWLTAACHVGPDKVLAALRVQLREQHLY-----YQDQLLPVSRI 107
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 127 EPHP-VYSWEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCHWISGWSIQD 185
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 IVHPQFYTAQIG--ADIALLEEFVNVSSHVHTVTLPPASSETFPFGMPGCVTWGSDVN 165
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 186 GVPLPHPTQKLKVPITDSEVCSHLYWRGAGQGP-----ITEDMLCAGYLEGERDACLDG 241
|||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 DERLPPFPPLKQVQKVPIMENHICDAKYLHGYTGDDVRIYRDDMLCAG--NTRDSCQGD 223
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 SGGPLMCQVDGAWLLAGIISWEGEACERNRPGVYISLSAHRWSVEKTV 289
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 224 SGGPLVCKVNGTWLQAGVSWEGEACQPNRPGIYTRVTVYLDWIHHVY 271
|||||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 9  
C35863



tryptase (EC 3.4.21.59) III precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 15-Jun-2001  
C:Accession: C35863; A38893  
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr  
A:Reference number: A35863; MUID:90251647  
A:Accession: C35863  
A:Molecule type: mRNA  
A:Residues: 9-275 <VAN>  
A:Accession: A35863  
A:Molecule type: DNA  
A:Residues: 1-9 <VA2>  
A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC03172.1; PID:g339977  
A:Note: The first nine residues of this sequence are inferred from genomic DNA of trypta  
R:Vanderslice, P.  
submitted to GenBank, April 1990  
A:Reference number: A38893  
A:Accession: A38893  
A:Molecule type: mRNA  
A:Residues: 9-131, 'K', 132-275 <VA3>  
A:Cross-references: GB:M33493; NID:g339984; PIDN:AAA36780.1; PID:g339985  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsin I #status predicted <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.4%; Score 522.5; DB 2; Length 275;  
Best Local Similarity 38.9%; Pred. No. 7.8e-37;  
Matches 112; Conservative 52; Mismatches 87; Indels 37; Gaps 9;  
QY 20 SLLLLASTAILNAARIPVPPACGKPOQLNRVVGEDSTDSEWPWIVSIQ---KNGTHCA 76  
Db 3 LLLLLALPVL--ASRAYAAPGQALQGVIGVGOEAPAPRSKPQVSLRVDRIYMHFCG 60  
QY 77 GSLTSRWVITAAHC-----FKDNLNKPFLFVLLGAWQLGNPGRSOKVGVAVW 126  
Db 61 GSLHPQWVLTAAHCVPDVKDLAALRVOLREOHLY-----YDQLLPVSR 107  
QY 127 EHPH-VYSWKEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGSIQD 185  
Db 108 IVHPQFYTAQIG--ADIALLELPVNVSSHVHTVLPAPSETFPPGMPCWVTCGWDVDN 165  
QY 186 GVPLPHTLOKLKVPIDSEVCSHLYWRGAGQP-----ITEDMLCAGYLEGERDACLDG 241  
Db 166 DERLPPPPPLKQVKVPMENHICDAKYLHGLAYTGDDVRIVRDDMLCAG--NTRDRSCGD 223  
QY 242 SGGPLMCOVDGAWLLAGIISWEGECAERNRPVYISLSAHSRSWVEKIV 289  
Db 224 SGGPLCKVNGTWLQAGVSVNGEGCAQPNRPGIYTRYTYILDWIIHHYV 271

RESULT 10  
S56160  
mast cell tryptase precursor - Mongolian jird  
C:Species: Meriones unguiculatus (Mongolian jird)  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 20-Jun-2000  
C:Accession: S56160  
R:Murakumo, Y.; Ide, H.; Itoh, H.; Tomita, M.; Kobayashi, T.; Maruyama, H.; Horii, Y.; N  
Biochem. J. 309, 921-926, 1995  
A:Title: Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil, Meriones u  
A:Reference number: S56160; MUID:95366971  
A:Accession: S56160  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-270 <MUR>  
A:Cross-references: EMBL:D31789; NID:g517122; PIDN:BA006598.1; PID:g517123  
C:Superfamily: trypsin; trypsin homology

F:26-262/Domain: trypsin homology <TRY>

Query Match 30.3%; Score 522; DB 2; Length 270;  
Best Local Similarity 38.0%; Pred. No. 8.5e-37;  
Matches 109; Conservative 52; Mismatches 84; Indels 42; Gaps 9;  
QY 21 LLLLASTAILNAARIPVPPACGKPOQLNRVVGEDSTDSEWPWIVSIQKNT---HHCAG 77  
Db 4 LLLLLALPFLSLMHSPLCQEWG-----IVGGOEAPGNKWPQVSLRANETVYRHFCCG 56  
QY 78 SLTSRWVITAAHC-----FKDNLNKPFLFVLLGAWQLGNPGRSOKVGVAVWE 127  
Db 57 SLHPQWVLTAAHCVPDVKDLAALRVOLRKQYLY-----YHDHLLAVSR 103  
QY 128 PHPV-YSWKEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGSIQDQ 186  
Db 104 THPTFYATQNG--ADIALLELPVNVSSHVHTVLPAPSETFPPGMPCWVTCGWDVDN 161  
QY 187 VPLPHTLOKLKVPIDSEVCSHLYWRGAGQP-----ITEDMLCAGYLEGERDACLDGS 242  
Db 162 VSLPPPPPLKEVQVNVENQICLKYHKGVYTGDNIIHIVRDDMLCAGN-EG-HDSCQGD 219  
QY 243 GGPLMCOVDGAWLLAGIISWEGECAERNRPVYISLSAHSRSWVEKIV 289  
Db 220 GGPLCKVNGTWLQAGVSVNGEGCALPNRPGIYTRYTYILDWIIHHYV 266

RESULT 11  
A47246  
tryptase (EC 3.4.21.59) 2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A47246  
R:McNeill, H.P.; Reynolds, D.S.; Schiller, V.; Childval, N.; Gurley, D.S.; Austen, K.F  
Proc. Natl. Acad. Sci. U.S.A. 89, 11174-11178, 1992  
A:Title: Isolation, characterization, and transcription of the gene encoding mouse ma  
A:Reference number: A47246; MUID:93087489  
A:Accession: A47246  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-273 <MCN>  
A:Cross-references: GB:L00653; NID:g200518; PIDN:AAA39992.1; PID:g200519  
A:Note: sequence extracted from NCBI backbone (NCBIN:119745, NCBIP:119746)  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:29-265/Domain: trypsin homology <TRY>

Query Match 30.1%; Score 518; DB 2; Length 273;  
Best Local Similarity 38.5%; Pred. No. 1.9e-36;  
Matches 110; Conservative 53; Mismatches 85; Indels 38; Gaps 9;  
QY 22 LLLASTAILNAARIPVPPACGKPOQLNRVVGEDSTDSEWPWIVSIQKNGT---HHCAGS 78  
Db 4 LLLLLTLLSSL---VHAAPGPMTRREGIVGGQAHGKNKWPQVSLRANETVYRHFCCGS 60  
QY 79 LLSRWVITAAHC-----FKDNLNKPFLFVLLGAWQLGNPGRSOKVGVAVWVEP 128  
Db 61 LIHPQWVLTAAHCVPDVKDLAALRVOLRKQYLY-----YHDHLMVTSQIT 107  
QY 129 HP-VYSWKEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGSIQDGV 187  
Db 108 HPDFYIVQDQ--ADIALKLTNPVNISDVHPVPLPAPSETFPPGMPCWVTCGWDINDGV 165  
QY 188 PLPHTLOKLKVPIDSEVCSHLYWRGAGQP-----ITEDMLCAGYLEGERDACLDGSG 243  
Db 166 NLPPPPPLKEVQVPLIENHICDLKYHKGLITGDNVHIVRDDMLCAGN-EG-HDSCQGD 223  
QY 244 GPLMCOVDGAWLLAGIISWEGECAERNRPVYISLSAHSRSWVEKIV 289  
Db 224 GPLVCKVEDTWLQAGVSVNGEGCAQPNRPGIYTRYTYILDWIIHHYV 269

RESULT 12  
 KOHUP  
 plasma kallikrein (EC 3.4.21.34) precursor - human  
 N:Alternate names: kininogenin; plasma prekallikrein  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 18-Jun-1999  
 C:Accession: A00921; A37939  
 R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.  
 Biochemistry 25, 2410-2417, 1986  
 A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four  
 A:Reference number: A00921; MUID:86243339  
 A:Accession: A00921  
 A:Molecule type: mRNA  
 A:Residues: 1-638 <CHU>  
 A:Cross-references: GB:M13143; NID:q190262; PIDN:AAA60153.1; PID:g190263  
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.  
 Biochemistry 30, 2050-2056, 1991  
 A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of  
 A:Reference number: A37939; MUID:91152016  
 A:Accession: A37939  
 A:Molecule type: protein  
 A:Residues: 20-27;40-46,'X',48,'H';50,'X',52-70,'H';75-76,'X',78-80;103-113;131-140;141-  
 260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X'  
 525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>  
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w  
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li  
 are linked by one or more disulfide bonds.  
 C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal  
 inogen and may also play a role in the renin-angiotensin system by converting prorenin i  
 C:Genetics:  
 A:Gene: GDB:KUK3  
 A:Cross-references: GDB:127575; OMIM:229000  
 A:Map position: 4q35-4q35  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-638/Product: plasma kallikrein heavy chain #status predicted <MAT>  
 F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>  
 F:20-109/Domain: apple repeat <AP1>  
 F:110-199/Domain: apple repeat <AP2>  
 F:200-289/Domain: apple repeat <AP3>  
 F:291-360/Domain: apple repeat <AP4>  
 F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>  
 F:391-621/Domain: trypsin homology <TRY>  
 F:21-104,47-77,51-57,111-134,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383  
 F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:318-347,340-345/Disulfide bonds: #status predicted  
 F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted  
 F:434,483,578/Active site: His, Asp, Ser. #status predicted

Db 571 KDACGDSGGPLVCKHNGMRLVGLTSGECARREQPVYTKVAEYMDWILEKTQSSD- 629  
 QY 295 RGRAQ 299  
 Db 630 -GRAQ 633  
 RESULT 13  
 KFHUI  
 coagulation factor XIa (EC 3.4.21.27) precursor [validated] - human  
 N:Alternate names: antihemophilic factor C; plasma thromboplastin antecedent  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Aug-1986 #sequence\_revision 26-May-1994 #text\_change 08-Dec-2000  
 C:Accession: A27431; A00920; A37940  
 R:Asakai, R.; Davie, E.W.; Chung, D.W.  
 Biochemistry 26, 7221-7228, 1987  
 A:Title: Organization of the gene for human factor XI.  
 A:Reference number: A27431; MUID:88107663  
 A:Accession: A27431  
 A:Molecule type: DNA  
 A:Residues: 1-625 <ASA>  
 A:Cross-references: GB:M18295  
 A:Note: the sequence shown follows the authors' translation  
 R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.  
 Biochemistry 25, 2417-2424, 1986  
 A:Title: Amino acid sequence of human factor XI, a blood coagulation factor with four  
 A:Reference number: A00920; MUID:86243360  
 A:Accession: A00920  
 A:Molecule type: mRNA  
 A:Residues: 1-625 <FUJ>  
 A:Cross-references: GB:M13142; NID:g182832; PIDN:AAA52487.1; PID:g182833  
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.  
 Biochemistry 30, 2056-2060, 1991  
 A:Title: Location of the disulfide bonds in human coagulation factor XI: the presence  
 A:Reference number: A37940; MUID:91152017  
 A:Accession: A37940  
 A:Molecule type: protein  
 A:Residues: 28-33;35-49,'X',51-55,'X',57-63;70-75,'X',77-79;107-109,'X',111-112;132-1  
 :280-282,'X',284;285-297;313-316,'X',318-319;320-326,'X',328-330,'X',347-349;373,'X',  
 C:Comment: The proenzyme consists of two identical chains linked by one or more disul  
 he active site, and a heavy chain, which associates with high molecular weight (HMW)  
 C:Genetics:  
 A:Gene: GDB:FLI1  
 A:Cross-references: GDB:119891; OMIM:264900  
 A:Map position: 4q35-4q35  
 A:Introns: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1;  
 C:Function:  
 A:Description: catalyzes the proteolytic activation of coagulation factor IX  
 A:Pathway: blood coagulation intrinsic pathway  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hy  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-387/Product: coagulation factor XIa heavy chain #status experimental <HCH>  
 F:19-108/Domain: apple repeat <AP1>  
 F:109-198/Domain: apple repeat <AP2>  
 F:199-288/Domain: apple repeat <AP3>  
 F:290-379/Domain: apple repeat <AP4>  
 F:388-625/Product: coagulation factor XIa light chain #status experimental <LCH>  
 F:388-618/Domain: trypsin homology <TRY>  
 F:20-103,514-581,571-599/Disulfide bonds: #status predicted  
 F:29/Disulfide bonds: interchain #status experimental  
 F:46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327  
 F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:339/Disulfide bonds: interchain #status predicted  
 F:387-388/Cleavage site: Arg-Ile (coagulation factor XIa) #status experimental  
 F:431,480,575/Active site: His, Asp, Ser #status predicted  
 F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 27.6%; Score 475.5; DB 1; Length 625;  
 Best Local Similarity 39.4%; Pred. No. 2e-32;  
 Matches 98; Conservative 39; Mismatches 97; Indels 15; Gaps 7;



[illegible]

Search completed: August 29, 2002, 10:24:52  
Job time: 656 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:24:12 ; Search time 13.46 Seconds  
(without alignments)  
911.894 Million cell updates/sec

Title: US-09-903-925A-263

Perfect score: 1720

Sequence: 1 MVVSGAPPALGGGCLGTFTS.....AGGGALRAPSGGGAARS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	1720	100.0	317	1 BSS4_HUMAN	Q9gzr4 homo sapien
2	1270.5	73.9	306	1 BSS4_MOUSE	Q9er10 mus musculus
3	739	43.0	290	1 MPN_HUMAN	Q9brg3 homo sapien
4	607.5	35.3	343	1 PSS8_HUMAN	Q16511 homo sapien
5	584	34.0	342	1 PSS8_RAT	Q9esb7 rattus norv
6	578.5	33.6	273	1 TRY1_SHEEP	Q9xsm2 ovis aries
7	574.5	33.4	342	1 PSS8_MOUSE	Q9esd1 mus musculus
8	569	33.1	321	1 TRY6_HUMAN	Q9nrr2 homo sapien
9	554	32.2	276	1 MCT6_MOUSE	P21845 mus musculus
10	546	31.7	275	1 TRY1_CANFA	P15944 canis famil
11	538.5	31.3	274	1 MCT6_RAT	P50343 rattus norv
12	532.5	31.0	275	1 TRY1_PIG	Q9n2d1 sus scrofa
13	531.5	30.9	314	1 TEST_HUMAN	Q9y6m0 homo sapien
14	530.5	30.8	275	1 TRB2_HUMAN	P20231 homo sapien
15	530	30.8	324	1 TEST_MOUSE	Q9jbi7 mus musculus
16	529.5	30.8	275	1 TRB1_HUMAN	Q15661 homo sapien
17	526.5	30.6	275	1 TRY1_HUMAN	P15157 homo sapien
18	522	30.3	270	1 TRY1_MERUN	P50342 meriones un
19	519.5	30.2	273	1 MCT7_RAT	P27435 rattus norv
20	519.5	30.2	311	1 TRY6_MOUSE	Q9qu17 mus musculus
21	518	30.1	273	1 MCT7_MOUSE	Q02844 mus musculus
22	490.5	28.5	455	1 TMS5_MOUSE	Q9er04 mus musculus
23	482.5	28.1	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
24	481	28.0	638	1 KAL1_HUMAN	P03952 homo sapien
25	475.5	27.6	625	1 FALL1_HUMAN	P03951 homo sapien
26	473	27.5	638	1 KAL_MOUSE	P26262 mus musculus
27	472.5	27.5	437	1 TMS4_HUMAN	Q9nrs4 homo sapien
28	472.5	27.5	454	1 TMS3_HUMAN	P57727 homo sapien
29	469.5	27.3	638	1 KAL_RAT	P14272 rattus norv
30	463	26.9	416	1 HEP5_MOUSE	Q35453 mus musculus
31	462	26.9	492	1 TMS2_HUMAN	O15393 homo sapien
32	460.5	26.8	812	1 PLMN_BOVIN	P06868 bos taurus
33	457.5	26.6	1034	1 ENTK_PIG	P98074 sus scrofa

ALIGNMENTS

RESULT 1					
ID	BSS4_HUMAN	STANDARD;	PRT;	317 AA.	
AC	Q9GZNR; 043342;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4)				
DE	(SP001LA).				
GN	PRSS26 OR PRSS22 OR BSSP4.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RA	Mitsui S., Okui A., Kominami K., Yamaguchi N.;				
RT	"Cloning and characterization of a human brain-specific serine				
RT	protease, hBSSP-4.;"				
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=Pancreas;				
RA	Wong G.W., Stevens R.L.;				
RT	"Identification of a new member of the chromosome 16 family of serine				
RT	proteases.;"				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RC	SEQUENCE OF 47-317 FROM N.A.				
RA	Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,				
RA	Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,				
RA	Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,				
RA	Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;				
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AB010779; BAB20263.1; -				
DR	EMBL; AF321182; AAG35070.1; -				
DR	EMBL; AC003965; AAB93671.1; -				
DR	MEROPS; S01.252; -				
DR	InterPro; IPR001314; Chymotrypsin.				
DR	InterPro; IPR001254; Trypsin.				
DR	Pfam; PF00089; trypsin; 1.				
DR	PRINTS; PR00722; CHYMOTRYPSIN.				
DR	SMART; SM00020; Tryp_SPC; 1.				
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.				

34	453.5	26.4	490	1 TMS2_MOUSE	Q9jig8 mus musculus
35	452	26.3	417	1 HEP5_HUMAN	P05981 homo sapien
36	451.5	26.2	855	1 ST14_MOUSE	P56677 mus musculus
37	450	26.2	416	1 HEP5_RAT	Q05511 rattus norv
38	450	26.2	810	1 PLMN_MACMU	P12545 macaca mula
39	446	25.9	338	1 PLMN_HORSE	P80010 equus cabal
40	446	25.9	790	1 PLMN_PIG	P06867 sus scrofa
41	446	25.9	810	1 PLMN_HUMAN	P00747 homo sapien
42	445	25.9	333	1 PLMN_CANFA	P80009 canis famil
43	445	25.9	786	1 STUB_DROME	Q05319 drosophila
44	444	25.8	1019	1 ENTK_HUMAN	P98073 homo sapien
45	443.5	25.8	1035	1 ENTK_BOVIN	P98072 bos taurus

```
DR PROSITE; PS00134; TRYPSIN_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN_SER; 1.  
KW Hydrolyase; Serine protease; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 317  
FT ACT_SITE 90 90  
FT ACT_SITE 141 141  
FT ACT_SITE 242 242  
FT ACT_SITE 242 242  
FT DISULFID 75 91  
FT DISULFID 175 248  
FT DISULFID 208 227  
FT DISULFID 238 266  
FT CARBOHYD 70 70  
FT CONFLICT 47 47  
SQ SEQUENCE 317 AA; E2A123BC86E79935 CRC64;  
  
Query Match 100.0%; Score 1720; DB 1; Length 317;  
Best Local Similarity 100.0%; Pred. No. 2.6e-141;  
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MVSGAPPALGGCLGFTSLLLAAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60  
DB 1 MVSGAPPALGGCLGFTSLLLAAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60  
  
QY 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAHCDFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120  
DB 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAHCDFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120  
  
QY 121 VGAWVEPHVPYVSKGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
DB 121 VGAWVEPHVPYVSKGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
  
QY 181 GSIDQGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240  
DB 181 GSIDQGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240  
  
QY 241 DSGGPLCMQVGDGAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300  
DB 241 DSGGPLCMQVGDGAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300  
  
QY 301 GGALRAPSGSGAAARS 317  
DB 301 GGALRAPSGSGAAARS 317  
  
RESULT 2  
ID BSS4_MOUSE STANDARD; PRT; 306 AA.  
AC Q9ER10;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4).  
GN PRSS26 OR BSSP4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Mitsui S., Okui A., Kominami K., Yamaguchi N.;  
RT "Cloning and characterization of a novel serine protease, mBSSP-4";  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AB010778; BAB20262.1; --  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN_SER; 1.  
KW Hydrolyase; Serine protease; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 306  
FT ACT_SITE 90 90  
FT ACT_SITE 141 141  
FT ACT_SITE 242 242  
FT DISULFID 75 91  
FT DISULFID 175 248  
FT DISULFID 208 227  
FT DISULFID 238 266  
FT CARBOHYD 70 70  
SQ SEQUENCE 306 AA; FBBF03C0C285E7E8 CRC64;  
  
Query Match 73.9%; Score 1270.5; DB 1; Length 306;  
Best Local Similarity 76.8%; Pred. No. 1.5e-102;  
Matches 232; Conservative 25; Mismatches 44; Indels 1; Gaps 1;  
  
QY 1 MVVSGAPPALGGCLGFTSLLLAAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60  
DB 1 MVVSGAPPALGGCLGFTSLLLAAILNAARIPVPACGKPKQQLNRVVGGEDSTDSE 60  
  
QY 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAHCDFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120  
DB 61 WPWIVSIQKNGTHRCAGSLTSTRWVITAHCDFKDNLNKPYLFSVLLGAWQLGNPGRSQK 120  
  
QY 121 VGAWVEPHVPYVSKGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
DB 121 VGAWVEPHVPYVSKGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
  
QY 181 GSIDQGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240  
DB 181 GSIDQGVPLPQTLQKLVPIIDSEVCSHLYWRGAGGPTEDMLCAGYLEGERDACL 240  
  
QY 241 DSGGPLCMQVGDGAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300  
DB 241 DSGGPLCMQVGDGAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIVQVQLRGAQG 300  
  
QY 301 GG 302  
DB 300 SG 301  
  
RESULT 3  
ID MPN_HUMAN STANDARD; PRT; 290 AA.  
AC Q9BQR3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Marapsin precursor (EC 3.4.21.-).  
GN MPN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;  
RT "Cloning, sequencing and expression of marapsin, a human serine  
RT proteinase.";
```

```

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ306593; CAC35467.1; -.
CC DR MEROPS: S01.074; -.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR InterPro: IPR001254; Trypsin.
CC DR Pfam: PF00089; trypsin; 1.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00020; Tryp_SPC; 1.
CC DR PROSITE: PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; zymogen; signal; glycoprotein.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT PROPEP 23 34 POTENTIAL.
CC FT CHAIN 35 290 MARAPLIN.
CC FT DOMAIN 35 277 SERINE PROTEASE.
CC FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 60 76 BY SIMILARITY.
CC FT DISULFID 158 235 BY SIMILARITY.
CC FT DISULFID 191 214 BY SIMILARITY.
CC FT DISULFID 225 253 BY SIMILARITY.
CC FT CARBOHYD 55 55 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 79 79 N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 290 AA; 31940 MW; 67BDC93BC70BEF7B CRC64;

Query Match 43.08; Score 739; DB 1; Length 290;
Best Local Similarity 48.88; Pred. No. 1.le-56;
Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;

QY 21 LLLASTAILNAARIPVPACGRQQLNRVVGSDTSEWPWIVSTQKNGTHHCASLL 80
DB 10 LLLLC-----FGSQRKAATACGRPMILNRMVGGDTQEGEPWQVSTQRNGSHFCGSLI 65

QY 81 TSWVITAACFKDNLNKPVLFSVLLGAWOLGNPGSRQKVGVAWEPHPVYSWKEGACA 140
DB 66 AEQWVLTAAHCFR-NTSETSLYQLLQARQLVQPGHAMYARVQVESNPPLYQ-GTASSA 123

QY 141 DIALVRLERSIQFSERVLPICLPDASTHLPPNTHCWTISGWSIQDGVPLPHPQTLQKLKV 200
DB 124 DVALVELEAPVPFTNYLTPLVCLPDPSVIFETGMNCVWTGWSPEEDLLPEPTLQKLAV 183

QY 201 PIIDSEVCSHLYWRGAGQ---PITEDMLCAGYLEGERDACIGDSGLPLCMQVDGAWLL 256
DB 184 PIIDTPKCNLLYKSDTEFGYQPKTKNDMLCAGFEKGKACKDGSGGLPVCLVCGQSWLQ 243

QY 257 AGIISWEGGAERNRPGVYISLSAHSRWKIVOGVLRGRAQGG 301
DB 244 AGVISWEGGACARQNPGRVYIRVTAAHHWIRHIIPLKLFQFARLGG 288

RESULT 4
PSS8_HUMAN STANDARD; PRT; 343 AA.
AC Q16651; Q9UCA3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Proctasin precursor (EC 3.4.21.-).
GN PRSS8.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC TISSUE-Prostate;
CC MEDLINE-95286644; PubMed-7768952;
CC Yu J.X., Chao L., Chao J.;
CC "Molecular cloning, tissue-specific expression, and cellular
CC localization of human proctasin mRNA.";
CC J. Biol. Chem. 270:13483-13489(1995).
CC [2]
CC SEQUENCE FROM N.A.
CC TISSUE-Placenta;
CC Strausberg R.;
CC Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC [3]
CC SEQUENCE OF 45-64.
CC TISSUE-Semen;
CC MEDLINE-94308140; PubMed-8034638;
CC Yu J.X., Chao L., Chao J.;
CC "Proctasin is a novel human serine proteinase from seminal fluid.
CC Purification, tissue distribution, and localization in prostate
CC gland.";
CC J. Biol. Chem. 269:18843-18848(1994).
CC -I- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.
CC -I- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A
CC DISULFIDE BOND.
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF
CC ITS C-TERMINUS.
CC -I- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,
CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR
CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL
CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL
CC FLUID.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L41351; AAC41759.1; -.
CC DR EMBL: U33446; AAB19071.1; -.
CC DR EMBL: BC001462; AA01462.1; -.
CC DR HSSP: P00763; IDPO.
CC DR MEROPS: S01.159; -.
CC DR MIM: 600823; -.
CC DR InterPro: IPR001314; Chymotrypsin.
CC DR InterPro: IPR001254; Trypsin.
CC DR Pfam: PF00089; trypsin; 1.
CC DR PRINTS: PR00722; CHYMOTRYPSIN.
CC DR SMART: SM00020; Tryp_SPC; 1.
CC DR PROSITE: PS00240; TRYPSIN_DOM; 1.
CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; zymogen; signal; glycoprotein;
CC Transmembrane.
CC FT SIGNAL 1 29 POTENTIAL.
CC FT PROPEP 30 32 ACTIVATION PEPTIDE.
CC FT CHAIN 33 44 PROCTASIN LIGHT CHAIN.
CC FT CHAIN 45 322 PROCTASIN HEAVY CHAIN.
CC FT PROPEP 323 343 POTENTIAL.
CC FT TRANSMEM 320 340 SERINE PROTEASE.
CC FT DOMAIN 45 286 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 37 154 BY SIMILARITY.
CC FT DISULFID 70 86 BY SIMILARITY.
CC FT DISULFID 168 244 BY SIMILARITY.

```

```
FT DISULFID 201 223 BY SIMILARITY.
FT DISULFID 234 262 BY SIMILARITY.
FT ACT_SITE 85 85 CHARGE RELAY SYSTEM.
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM.
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM.
FT CARBOHYD 159 159 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 343 AA; 36431 MW; 98DD6447F5A8C1B2 CRC64;

Query Match 35.3%; Score 607.5; DB 1; Length 343;
Best Local Similarity 44.2%; Pred. No. 2.9e-45;
Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGCLCTFTSL---LLASTAILNAARI--PVPPACKGPQOLNRVVGGE--STSEWPIVSIKNG 66
DB 7 LGPQLGAVAILLYGLLRSTGAGAEAP-----CGVAPQA-RITGSSAVAGOWPQVVS 61

QY 67 IQKNGTHHCAGSLTTSRWVITAAHCFKDNLPKLYFVLLGAWQLGNPGRSQKVGVAWVPHV 126
DB 62 ITYEGVHVCGLSVSEQWLSAAHCFPEHKE-AVEVKLGALHQLDSYSEDAKVSTLKI 120

QY 127 EPHVYSWKECACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCHWISGWSIQDG 186
DB 121 IPHPSY-LQESQGDIALQLSRPITFSRIRPCLPAANASFPNGLUHCTVTGWHVAPS 179

QY 187 VPLPHPQTLQKLVPIIDSEVCSHLYWRGA--GQGP--ITEDMLCAGYLEGERDACLGDGSG 243
DB 180 VSLTPRPLQQLVPLISRETCLNYLNIDAKPEHPHFQEDMVCAGIVEGGKDACQCGDSG 239

QY 244 GPLMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVQGVQLR 295
DB 240 GPLSCPVEGLWYLTGIVSWGDACGARNRPGVYTLASSYASWISQKVTLEQPR 291

RESULT 5
PSS8_RAT STANDARD; PRT; 342 AA.
AC Q9ES87; Q9ER01;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Prostatin precursor (EC 3.4.21.-).
GN PRS8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RC Tissue=Kidney;
RA Adachi M., Kitamura K., Miyoshi T., Tomita K.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.

Wang C.;
RT "Molecular cloning and expression of rat prostatin.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSESSES A TRYPsin-LIKE CLEAVAGE SPECIFICITY (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF ITS C-TERMINUS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
CC EMBL; AB017638; BAB20281.1; -.
DR EMBL; AF202076; AAG32641.1; -.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Glycoprotein;
KW Transmembrane.
FT SIGNAL 1 29 POTENTIAL.
FT PROPEP 30 32 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 33 44 PROSTATIN LIGHT CHAIN.
FT CHAIN 45 322 PROSTATIN HEAVY CHAIN.
FT PROPEP 323 342 BY SIMILARITY.
FT TRANSMEM 320 340 POTENTIAL.
FT DOMAIN 45 286 SERINE PROTEASE.
FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).
FT DISULFID 70 86 BY SIMILARITY.
FT DISULFID 168 244 BY SIMILARITY.
FT DISULFID 201 223 BY SIMILARITY.
FT DISULFID 234 262 BY SIMILARITY.
FT ACT_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 159 159 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 17 17 I -> V (IN REF. 1).
FT CONFLICT 292 292 A -> V (IN REF. 1).
SQ SEQUENCE 342 AA; 36843 MW; 5EDIAF05D9213898 CRC64;

Query Match 34.0%; Score 584; DB 1; Length 342;
Best Local Similarity 43.0%; Pred. No. 3e-43;
Matches 125; Conservative 43; Mismatches 113; Indels 10; Gaps 7;

QY 15 LGTFTSLLLASTAILNAARI---PVPPACKGPQOLNRVVGGE--STSEWPIVSIKNG 71
DB 9 LGQLEALFILLGLLQ-SRIGADGTREASCGAVIQ-PRITGGGSAKPGQWQVSIYNG 66

QY 72 THHCAGSLTTSRWVITAAHCFKDNLPKLYFVLLGAWQLGNPGRSQKVGVAWVPHV 131
DB 67 VHVCGSLGSNQQWVVSAAHCFPEHKEE-YEVKLGALHQLDSFNSDIIVHTVAQIISSH 125

QY 132 YSWKEGACADIALVRLERSIQFSERVLPICLPDASHLPPNTHCHWISGWSIQDGVLPH 191
DB 126 YR-EEGSGQDIALRLSSPVTFSRYIRPCLPAANASFPNGLUHCTVTGWHVAPSVSLQT 184

QY 192 PQTLQKLVPIIDSEVCSHLY-WRGAQGP--ITEDMLCAGYLEGERDACLGDGSGPLMC 248
DB 185 PRPLQQLVPLISRETCLNYLNIDAKPEHPHFQEDMVCAGIVEGGKDACQCGDSG 244

QY 249 QVPGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVQGVQLRGAQ 299
DB 245 PIDGLWYLAGIVSGDAGCAPNRPVYTLSTYASWIIHHVAELQPRVQ 295

RESULT 6
TRYP_SHEEP
ID TRYP_SHEEP STANDARD; PRT; 273 AA.
AC Q9XSW2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin 2 precursor (EC 3.4.21.59).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
[1]
RN SEQUENCE FROM N.A.
RC Tissue=Abomasum;
RA Pemberton A.D., McAleese S.M., Huntley J.F., Mackellar A.,
RA Collier D.D.S., McMillan L., Scudamore C.L., Miller H.R.P.;
```



```

"DNA sequence of sheep mast cell tryptase and its immunolocalisation
in lung, skin and gut in comparison with sheep mast cell proteinase-
1.";
Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION
RESPONSE OF THIS CELL TYPE.
-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but
with more restricted specificity than trypsin.
-!- SUBUNIT: HOMOTETRAMER (By similarity).
-!- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON
MAST CELL ACTIVATION.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Y18224; CAB41989.1; -.
HSSP: P20231; 1AAO.
MEROPS; S01.143; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP-SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 28 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 29 273 TRYPTASE 2.
FT ACT_SITE 72 72 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 57 73 BY SIMILARITY.
FT DISULFID 153 228 BY SIMILARITY.
FT DISULFID 186 209 BY SIMILARITY.
FT DISULFID 218 246 BY SIMILARITY.
FT CARBOHYD 231 231 N-LINKED (GLCNAC.. ) (POTENTIAL).
SQ SEQUENCE 273 AA; 30288 MW; DE9BA79218C3E67D CRC64;
Query Match 33.6%; Score 578.5; DB 1; Length 273;
Best Local Similarity 41.7%; Pred. No. 7.1e-43;
Matches 115; Conservative 53; Mismatches 93; Indels 15; Gaps 7;
QY 21 LLLLAATAINAAARIPIPPAGCKGKQQLNRRVVGSDSTDSEWPWIVSIQ---KNGTTHHCAG 77
| | | | : : : : | | | : : : : : | | | | : : : : |
Db 2 LHLALALLLSL--VSRAAPGAQLQRSGIIGKEAPCSRWPQWSLRVDQYWRHOCGG 59
| | | | : : : : | | | : : : : : | | | | : : : : |
QY 78 SLTSRWVITAAHCFKONLNKPYLFLSVLLGAWOLGNPGSRSKYGVAVPEHPHYYSWKEG 137
| | : : | | | | : : : : | | : : : : : : : : | | : : |
Db 60 SLIHQPWVLTAAHCHIGELOEPSDFRVQLREOHL---YYQRLLELPISRVIIPHHYVVEN 116
| | | | : : : : | | | : : : : : | | | | : : : : |
QY 138 ACADIALVRLESITQSFRVLPICLPDASITHLPNTHCWISGWGSIODGPVLPHPQTLOK 197
| | | | : : : : | | | : : : : : | | | | : : : : |
Db 117 G-ADIALLQLLEEPVISRHRVOPVTLPPESETFPPESCQCVTWGWDVNGRPLPPPYPLKQ 175
| | | | : : : : | | | : : : : : | | | | : : : : |
QY 198 LKVPIDISEVCSHLYWRGAGQG---PIT-EMLCAKYLEGERDACLGDSGGLPMQCDGA 253
| | | | : : : : | | | : : : : : | | | | : : : : |
Db 176 VKVPDIVNSCDWKYHSGLSLDYSVIPVEDNLCAAG--DGRDSCQGDGSLPKVNGT 233
| | | | : : : : | | | : : : : : | | | | : : : : |
QY 254 WLAGIISWGEGCAERNRPBGVYISLSAHRSWVEKIV 289
| | | | : : : : | | | : : : : : | | | | : : : : |
Db 234 WLQAGVYSWGDCGAKPNRPGLYTRITSYLDWIHQYV 269
| | | | : : : : | | | : : : : : | | | | : : : : |

```

Db 7 LGLQLEAVTILLILG--LLOSGIRADGTEASCGAVIO-PRITGGGSAKPQWPNQWISIT 63

Qy 69 KNGTHHCAGSLTTRWVITAAHCF-KDNLNKPYLEFVLLGAWQLGNPGSRQKVGVAWVE 127

Db 64 YDGNHVCGLSVNKNWVYSAHCFPREHSREAY--EVKLGAHQLDYSNDIVVHTVAQII 121

Qy 128 PHPVYSWKEGACADIALVRLERSIOFSEVLPICLPDASIHLPNTHCWISGWSIODGV 187

Db 122 THSSYR-EEGSGDIAFRLSPVTFYSYIRPICLPAANAFPNGLCHCTVTGWHVAPSV 180

Qy 188 PLHPOTLQKLVPIIDSEVSHLY-WRAGAGQP--ITEDMLCAGYLEGERDACIGDSGG 244

Db 181 SLQTPRLOQLEVEPISEITSCSYLNINAVPEPHTIQDDMLCAGYVKGKDACQGDGG 240

Qy 245 PLMCQVDGAWLLAGIISWGECAERNRPGVYISLSAHRSWVEKIVQGVQLR 295

Db 241 PLSCPMEGIWLAGIVSGWDACGAPNRPVYTLFTSTVASHHHVHVAELQPR 291

RESULT 8

TRYG\_HUMAN STANDARD; PRT; 321 AA.

AC Q9NR2; Q9NRQ8; Q9C015; Q9UBB2; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Tryptase gamma precursor (EC 3.4.21.-) (Transmembrane tryptase).

GN TP5G1 OR TMT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).

RX MEDLINE=20302813; PubMed=10843716;

RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallao M., Walters P.J., Verghese G.M.;

RT "Characterization of human gamma-tryptases, novel members of the chromosome 16p mast cell tryptase and prostatic gene families.";

RL J. Immunol. 164:6566-6575(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99452974; PubMed=10521469;

RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C., Friend D.S., Krills S.A., Stevens R.L.;

RT "Identification of a new member of the tryptase family of mouse and human mast cell proteases which possesses a novel COOH-terminal hydrophobic extension.";

RL J. Biol. Chem. 274:30784-30793(1999).

RN [3]

RP SEQUENCE OF 220-321 FROM N.A.

RA Mittman S., Agnew W.S.;

RT "Organization and alternative splicing of CACNA1H.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (Potential).

CC -1- TISSUE SPECIFICITY: Expressed in many tissues.

CC -1- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which differ by 5 residues.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. TRYPTASE SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF191031; AAF76457.1; -

DR EMBL; AF195508; AAF76458.1; -

DR EMBL; AF175759; AAF03697.1; -

DR EMBL; AF175522; AAF03695.1; -

DR EMBL; AF2233563; AAG48852.2; -

DR MEROPS; S01.028; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; FALSE\_NEG.

KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;

KW Transmembrane; Polymorphism.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.

FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.

FT TRANSMEM 284 304 POTENTIAL.

FT ACT\_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 26 145 INTERCHAIN (POTENTIAL).

FT DISULFID 63 79 BY SIMILARITY.

FT DISULFID 159 228 BY SIMILARITY.

FT DISULFID 192 210 BY SIMILARITY.

FT DISULFID 218 246 BY SIMILARITY.

FT CARBOHYD 85 85 N-LINKED (GLCNAC... (POTENTIAL).

FT VARIANT 60 60 /FTIG=VAR\_012097.

FT VARIANT 126 126 I -> M (IN GAMMA-II).

FT VARIANT 132 132 S -> T (IN GAMMA-II).

FT VARIANT 204 204 /FTID=VAR\_012099.

FT VARIANT 288 288 L -> I (IN GAMMA-II).

FT VARIANT 288 288 /FTID=VAR\_012100.

FT CONFLICT 160 160 L -> F (IN GAMMA-II).

FT CONFLICT 160 160 /FTID=VAR\_012101.

FT CONFLICT 160 160 W -> S (IN REF. 1).

SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;

Query Match 33.1%; Score 569; DB 1; Length 321;

Best Local Similarity 41.1%; Pred.No. 5.6e-42;

Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

Qy 15 LGTFTSLLLASTAILNAARIPVPPACGKPKQLN---RVVGGEDSTSEWPWISIQNG 71

Db 3 LGACGLLLLL---AVPGVSLRTLPQCGRPQVSDAGGIRVGGHAAPAGANPWQASLR 59

Qy 72 THHCAGSLTTRWVITAAHCFKDNLPYLFVLLGAWQLGNPGSRQKVGVAWVEHPV 131

Db 60 MHVCGGSLSPQWVLTAAHCFSGSLSSD-YQVHLGEITLSPHFST---VRQILHSS 115

Qy 132 YSWKEGACADIALVRLERSIOFSEVLPICLPDASIHLPNTHCWISGWSIODGVPLPH 191

Db 116 PSQPGTSGDIALVELSVPTLSRILPVCLPEASDDFCGRCWVGTGWTREGEPLPP 175

Qy 192 POTLQKLVPIIDSEVSHLYWRAGAGQPITEDMLCAGYLEGERDACIGDSGGLMCOVD 251

Db 176 PYSLEKVSVDVETCTRRDY-FPGGSIQLPDMICA---RPGCDACQDSDSGGLVCOVN 231

Qy 252 GAWLLAGIISWGECAERNRPGVYISLSAHRSWVEKIVQGVQLRGAQGG---GALRAP 307

Db 232 GAWVQAGISWGECCGRPNRPGVYTRPVYVNNIRRI-----TASGGSEGYPLRP 283

RESULT 9

MCT6\_MOUSE STANDARD; PRT; 276 AA.

ID MCT6\_MOUSE AC P21845; Q61962;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mast cell protease 6 precursor (EC 3.4.21.59) (MMP-6) (Tryptase).

GN MCTP6.

OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91139682; PubMed=1995638;  
 RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;  
 RT "Cloning of the cDNA and gene of mouse mast cell protease-6.  
 RT Transcription by progenitor mast cells and mast cells of the  
 RT connective tissue subclass.";  
 RL J. Biol. Chem. 266:3847-3853(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=LEADEN X A1;  
 RA Huang R., Abrink M., Gobl A.E., Nilsson G., Aveskogh M., Larsson L.G.,  
 RA Nilsson K., Hellman L.;  
 RT "Expression of a mast cell tryptase in the human monocytic cell lines  
 RT U-937 and Mono Mac 6.";  
 RL Scand. J. Immunol. 38:359-367(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RX STRAIN=LEADEN X A1;  
 RA MEDLINE=95048582; PubMed=7959952;  
 RA Huang R., Hellman L.T.;  
 RT "Genes for mast-cell serine protease and their molecular evolution.";  
 RL Immunogenetics 40:397-414(1994).  
 RN [4]  
 RP SEQUENCE OF 32-54.  
 RX MEDLINE=90222202; PubMed=2326280;  
 RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
 RA Serafin W.E.;  
 RT "Different mouse mast cell populations express various combinations  
 RT of at least six distinct mast cell serine proteases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF MCPT6 ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SHORT FORM IS PROBABLY  
 CC A NON FUNCTIONAL VARIANT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M57626; AAA39988.1; -;  
 DR EMBL; M57625; AAA39987.1; -;  
 DR EMBL; L31853; AAA39725.1; -;  
 DR EMBL; X78542; CAA55288.1; -;  
 DR PIR; A38654; A38654.  
 DR PIR; D35646; D35646.  
 DR HSSP; P20231; 1AAO.  
 DR MEROPS; S01.025; -;  
 DR MGD; MGI:96942; Mcpt6.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00089; trypsin.1  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;

KW Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 31  
 FT CHAIN 32 276  
 FT ACT\_SITE 75 75  
 FT ACT\_SITE 122 122  
 FT ACT\_SITE 225 225  
 FT ACT\_SITE 225 225  
 FT DISULFID 60 76  
 FT DISULFID 156 231  
 FT DISULFID 189 212  
 FT DISULFID 221 249  
 FT CARBOHYD 133 133  
 FT VARSPPLIC 223 230  
 FT VARSPPLIC 231 276  
 SQ SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;  
 Query Match 32.2%; Score 554; DB 1; Length 276;  
 Best Local Similarity 41.6%; Pred. No. 9.4e-41;  
 Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;  
 Qy 21 LLLASTAILMAARIPVP-PACGKPPQQLNRVVVGSDSDSEWPWIVSIQ---KNGTHHCA 76  
 Db 6 LLLWALSLLASLVYSAPRPA---NQRVGVGGHEASESKWPQVSLRFLNLYWIHFCG 61  
 Qy 77 GSLTSTRWVITAAHCCKDNLNKPYLSVLLGAWQL--GNPGRSOKVGVAVWVPEHPVYSW 134  
 Db 62 GSLIHQPWLVTAACHCVGPHIKSPQLFRVQLRQYLYYGD-----QLLSLNRIIVVHPHYT 116  
 Qy 135 KEGACADIALYRLERSIOFSERVLPICLPDASIHLPNTHCWISGWGSIODGVPLPHPQT 194  
 Db 117 AEGG-ADVALLEVPVNVNTHIPIPSLPASETFPPGTSCWVTGWGDINDPELPPIYP 175  
 Qy 195 LQKLVPIIDSEVCSHLYWRGAGQG---PITED-MLCAGYLEGERDADCLGDSGGPLMQV 250  
 Db 176 LKQVKPIVENSCLDKRYHTGLTGDGDFIVHDGMLCAG--NTRDSCGDSGGPLVKV 233  
 Qy 251 DGALLAGIISWEGCAERNRPGVVISLSAHSRWEKIV 289  
 Db 234 KGTWLQAGVSWEGCAQPNKPGITRYTYLYLDWIHYV 272  
 RESULT 10  
 TRYP\_CANFA  
 ID TRYP\_CANFA STANDARD; PRT; 275 AA.  
 AC P15944;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tryptase precursor (EC 3.4.21.59).  
 OS Canis familiaris (dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=49352460; PubMed=2504277;  
 RA Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;  
 RT "Molecular cloning of dog mast cell tryptase and a related protease:  
 RT structural evidence of a unique mode of serine protease activation.";  
 RL Biochemistry 28:4148-4155(1989).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
 CC MAST CELL ACTIVATION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



Db 58 CGSLHPQWLVTAACVGLHIKSPFLRVQL-----REQLYYADQLLTYNRTV 107

QY 126 VEPHYYSKMGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWSIQD 185

Db 108 VHPH-VYTVEDG--ADIALLELEIPNVSTHPIPLPASETFFSGTSCWWTGWGDIDS 164

QY 186 GYPLPHQPLQKLVPIIDSEVCSHLYWRGAGQ---PITED-MCAGYLEGERDACLGD 241

Db 165 DEPLPYPPLKQVKVPIVENSCLDRYHTGLYTGDDVPVQDGMCLAG--NTRSDSCQGD 222

QY 242 SGPLMCOVDGAWLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIV 289

Db 223 SGGLPVCKVGTWLQAGVSWGEGCAERNRPGIYTRVYLDWIHRIV 270

RESULT 12

TRYPT\_PIG STANDARD; PRT; 275 AA.

AC Q9N2D1;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Trypsin precursor (EC 3.4.21.59).

GN MCT7.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI\_TaxId=9823;

RN [1]

RP TISSUE=Lung;

RA Chen Y., Shiota M., Ohuchi M., Towatari T., Tashiro J., Murakami M.,

RA Yano M., Yang B., Kido H.;

RT "A novel trypsin-type protease from pig lungs, triggering infection by

RT pneumotropic viruses: purification and characterization."

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST

CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION

CC RESPONSE OF THIS CELL TYPE.

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but

CC with more restricted specificity than trypsin.

CC -1- SUBUNIT: HOMOTETRAMER (By similarity).

CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON

CC MAST CELL ACTIVATION.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; AB038652; BAA93614.1; -

DR MEROPS; S01.143; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYPSIN\_SPC; 1.

DR PROSITE; PS02440; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Hydrolase; Serine protease; Signal; Glycoprotein; zymogen.

FT SIGNAL 1 20 POTENTIAL.

FT PROPP 21 30 ACTIVATION PEPTIDE (BY SIMILARITY).

FT CHAIN 31 275 TRYPTASE.

FT ACT\_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 59 75 BY SIMILARITY.

FT DISULFID 155 230 BY SIMILARITY.

FT DISULFID 188 211 BY SIMILARITY.

FT DISULFID 220 248 BY SIMILARITY.

FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 275 AA; 30439 MW; ACC582647FCCB973 CRC64;

Query Match 31.0%; Score 532.5; DB 1; Length 275;

Best Local Similarity 39.4%; Pred. No. 6.7e-39;

Matches 109; Conservative 53; Mismatches 100; Indels 15; Gaps 7;

QY 20 SLLLLASTAILNAARIPVPVACPKPOOLNRVVGEDSTSENPWISIQ--KNGTHCA 76

Db 3 NLLVLAFLPLVSL--VHTAPAPQALERAGIVGKAPGHKWPQVSLRCLDQYWKHFCG 60

QY 77 GSLLTSRWVITAAHCFKDNLNKPYLFSLVGLGAMQLNPGSRKQVGVAVWEPHPVYSWKE 136

Db 61 GSLLHFWQWLTAACFCGPEKADPLYIRVQLGEOHL---YYQDRLLLVSRIVHPNY-YDE 116

QY 137 GACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWSIQDGVPLPHPQTLQ 196

Db 117 VNCADIALLEDPVNLSSHVQPVTLPPASETFFKGTFCWVTGWDVHSGWPLPPVPLK 176

QY 197 KLVPIIDSEVCSHLYWRGAGQ---ITEDMLCAGYLEGERDACLGDGSGPLMCOVDG 252

Db 177 QVRVPIVENSECDMQYHLGLSTGDNPIVRDMLCAG-SEG-HDSCQDGGSLVCRVNG 234

QY 253 AMLLAGIISWGEGCAERNRPGVYISLSAHRSWVEKIV 289

Db 235 TWLQAGVSWGEGCALPFRPGIYTRVYLDWIHQCI 271

RESULT 13

TEST\_HUMAN STANDARD; PRT; 314 AA.

AC Q9Y6M0; Q9NS34; Q9P2V6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Testisin precursor (EC 3.4.21.-) (Eosinophil serine protease 1) (ESP-

DE 1).

GN PRS21 OR TEST1 OR ESPL.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Eosinophil;

RX MEDLINE=99045401; PubMed=9826525;

RA Inoue M., Kanbe N., Kurosawa M., Kido H.;

RT "Cloning and tissue distribution of a novel serine protease esp-1 from

RT human eosinophils."

RL Biochem. Biophys. Res. Commun. 252:307-312(1998).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RX MEDLINE=20068805; PubMed=10600542;

RA Inoue M., Isobe M., Itoyama T., Kido H.;

RT "Structural analysis of esp-1 gene (PRSS 21)."

RL Biochem. Biophys. Res. Commun. 266:564-568(1999).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Cervical carcinoma;

RX MEDLINE=99323395; PubMed=10397266;

RA Hooper J.D., Nicol D.L., Dickinson J.L., Eyre H.J., Scarman A.L.,

RA Normyle J.F., Stuttgart M.A., Douglas M.L., Loveland K.A.,

RA Sutherland G.R., Antalis T.M.;

RT "Testisin, a new human serine proteinase expressed by premeiotic

RT testicular germ cells and lost in testicular germ cell tumors."

RL Cancer Res. 59:3199-3205(1999).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE-20461760; PubMed-11004480;  
 RA Hooper J.D., Bowen N., Marshall H., Cullen L.M., Sood R., Daniels R.,  
 RA Stuttgart M.A., Normyle J.F., Higgs D.R., Kastner D.L., Ogbourne S.M.,  
 RA Perra M.F., Jazwinska E.C., Antalio T.M.;  
 RT "Localization, expression and genomic structure of the gene encoding  
 the human serine protease testisin.";  
 RL Biochim. Biophys. Acta 1492:63-71(2000).  
 CC -!- FUNCTION: COULD REGULATE PROTEOLYTIC EVENTS ASSOCIATED WITH  
 TESTICULAR GERM CELL MATURATION.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (potential).  
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/L (SHOWN HERE), 2/S AND 3; ARE  
 PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PREMEIOTIC  
 TESTICULAR GERM CELLS, MOSTLY LATE PACHYTENE AND DIPTOTENE  
 SPERMATOCYTES.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC  
 DR EMBL; AF058300; AAD41588.1; -;  
 DR EMBL; AB031329; BAA83520.1; -;  
 DR EMBL; AB031330; BAA83521.1; -;  
 DR EMBL; AB031331; BAA89532.1; -;  
 DR EMBL; AF058301; AAF79019.1; -;  
 DR EMBL; AF058301; AAF79020.1; -;  
 DR HSP; P20231; 1AAG.  
 DR MEROPS; S01.011; -;  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Trypsin.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPSIN\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Signal; GPI-anchor;  
 KW zymogen; Alternative splicing.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT PROPEP 20 41 POTENTIAL.  
 FT CHAIN 42 288 POTENTIAL.  
 FT PROPEP 289 314 REMOVED IN MATURE FORM (POTENTIAL).  
 FT ACT\_SITE 82 82 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 137 137 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT DISULFID 33 157 POTENTIAL.  
 FT DISULFID 67 83 POTENTIAL.  
 FT DISULFID 171 244 POTENTIAL.  
 FT DISULFID 204 223 POTENTIAL.  
 FT DISULFID 234 262 POTENTIAL.  
 FT LIPID 288 288 GPI-ANCHOR (POTENTIAL).  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 200 200 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 273 273 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPPLIC 87 88 MISSING (IN ISOFORM 2).  
 FT VARSPPLIC 222 235 MISSING (IN ISOFORM 3).  
 SQ SEQUENCE 314 AA; 34884 MW; E738CF73F6B56E98 CRC64;  
 Query Match . 30.9%; Score 531.5; DB 1; Length 314;  
 Best Local Similarity 36.0%; Pred. No. 9.4e-39;  
 Matches 111; Conservative 57; Mismatches 115; Indels 25; Gaps 6;  
 QY 13 GCGTFTSLTSLASTAIL---NAARIPVPPACGKQQLNRRVGGEDSTDSEMPWIVSIQK 69  
 DB 2 GARGALLALLARACLRKPESQEAAPLSGPGRRVITSIRVGGEDALGRWPQGSRLR 61

QY 70 NGTHHCAGSLTTRWVITAAHCFK--DNLNKPYLFSVLLGA-----WOLGNPGRSQK 120  
 DB 62 WDSHVGCVSLTSHRWALTAHCFETYSDLSPSGWVQFGLTSMPSFWSLQAYTR-YF 120  
 QY 121 VGVAVPEPHVPSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
 DB 121 VSNIIYLSR-----YLGNSPYDIALVLSAPVTYTKHIQICLQASTFEFFENRDCWWTGW 176  
 QY 181 GSTQDGVPLPHPTLOKLVPIIDSEVCSHLYWRGAGOGPITEDMLCAGYLEGERDACL 240  
 DB 177 GYKDEALPSPHTLOEVQVAILNNSMCHFLKYSFRKDFGDMVWAGNAQGGKACFG 236  
 QY 241 DSGGLQMLQVQDGAWLLAGIISWEGEACERNRPGVYISLSAHRSVKRVGVQLRGRAG 300  
 DB 237 DSGGLACNKNGLWYQIGVSVGCGCRPNRPGVYINSHHFEWIKLM-----AQS 288  
 QY 301 GGALRAPPS 308  
 DB 289 GMSQDPDS 296  
 RESULT 14  
 TRB2\_HUMAN  
 ID TRB2\_HUMAN STANDARD; PRT; 275 AA.  
 AC P20231; Q9UQI7; Q9UQI6; Q15664; O95827;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin beta-2 precursor (EC 3.4.21.59) (Trypsin 2) (Trypsin II).  
 GN TPB2 OR TPB2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
 RC TISSUE=Lung;  
 RX MEDLINE=90369005; PubMed=2203827;  
 RA Miller J.S., Moxley G., Schwartz L.B.;  
 RT "Cloning and characterization of a second complementary DNA for human  
 trypsin.";  
 RL J. Clin. Invest. 86:864-870(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
 RX MEDLINE=93166209; PubMed=8434231;  
 RA Blom T., Hellman L.;  
 RT "Characterization of a trypsin mRNA expressed in the human basophil  
 cell line K0812.";  
 RN [3]  
 RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
 RX MEDLINE=90251647; PubMed=2187193;  
 RA Vanderslice P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S.,  
 RA Caughey G.H.;  
 RT "Human mast cell trypsin: multiple cDNAs and genes reveal a  
 multigene serine protease family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
 RX MEDLINE=99121069; PubMed=9920877;  
 RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
 RT "Characterization of genes encoding known and novel human mast cell  
 trypsinases on chromosome 16p13.3.";  
 RL J. Biol. Chem. 274:3355-3362(1999).  
 RN [5]  
 RP REVISIONS.  
 RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).  
 RX MEDLINE=98180625; PubMed=9521329;  
 RA Pereira P.J.B., Berghner A., Macedo-Ribeiro S., Huber R.,

Matschner G., Fritz H., Sommerhoff C.P., Bode W.;  
"human beta-tryptase is a ring-like tetramer with active sites facing  
a central pore.";  
Nature 392:306-311(1998).  
[7]  
X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).  
MEDLINE-99432168; PubMed-10500112;  
Sommerhoff C.P., Bode W., Pereira P.J.B., Stubbs M.T.,  
Stuerzebecher J., Piechotka G.P., Matschner G., Bergner A.;  
"The structure of the human beta1-tryptase tetramer: fo(u)r better or  
worse.";  
Proc. Natl. Acad. Sci. U.S.A. 96:10984-10991(1999).  
-1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
RESPONSE OF THIS CELL TYPE.  
-1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|- , but  
with more restricted specificity than trypsin.  
-1- SUBUNIT: HOMOTETRAMER.  
-1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
MAST CELL ACTIVATION.  
-1- POLYMORPHISM: There are two alleles; beta-II and beta-III which  
differ by 3 residues.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY. TRYPTASE SUBFAMILY.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M37488; AAA51843.1; -  
EMBL; M33492; AAA36779.1; -  
EMBL; M33493; AAA36780.1; -  
EMBL; S55531; AAD13876.1; -  
EMBL; AF099143; AAD17859.2; -  
EMBL; AF099145; AAD17857.1; -  
EMBL; AF099146; AAD17858.1; -  
PIR; A37193; A37193.  
PIR; B35863; B35863.  
PDB; IAO; 31-JAN-94.  
PDB; IAO; 23-MAR-99.  
MEROPS; S01.027; -  
InterPro; IPR001314; Chymotrypsin.  
InterPro; IPR001254; Trypsin.  
Pfam; PF00089; trypsin; 1.  
PRINTS; PR00722; CHYMOTRYPSIN.  
SMART; SM00020; TRYPSIN\_DOM; 1.  
PROSITE; PS00240; TRYPSIN\_DOM; 1.  
PROSITE; PS00134; TRYPSIN\_HIS; 1.  
PROSITE; PS00135; TRYPSIN\_SER; 1.  
Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;  
Polymorphism; 3d-structure.  
SIGNATURE  
PROPEP 19 30  
CHAIN 31 275  
ACT\_SITE 74 74  
ACT\_SITE 121 121  
ACT\_SITE 224 224  
DISULFID 59 75  
DISULFID 155 230  
DISULFID 188 211  
DISULFID 220 248  
CARBOHYD 233 233  
VARIANT 51 53  
POTENTIAL.  
ACTIVATION PEPTIDE.  
TRYPTASE BETA-2.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
N-LINKED (GLCNAC... ) (POTENTIAL).  
HGP -> RDR (IN BETA-III).  
/FTIG=VAR\_012104.  
K -> N (IN REF. 3).  
CONFLICT 132 132  
SEQUENCE 275 AA; 30529 MW; 2B27396C51F5C7A0 CRC64;  
Query Match 30.8%; Score 530.5; DB 1; Length 275;

Best Local Similarity    39.2%;    Pred. No. le-38;  
Matches 113; Conservative    53; Mismatches    85; Indels    37; Gaps    9;

QY    20 SLILLASTAILNAAIRVPPACGRRPOOLNRVVGGEDSDTSSEWPTVTSTOKNG---THHCA 76  
         :|||:|||:: |||:: |::| ::|||::|::|::|::|::|::|::|  
Db     3 NLLLALPVL--ASRAYAAPAGLOARVGIVGGCEAPRSKWPQVSURVHGPTWMHFPG 60  
  
QY    77 GSLSLRWTAAHC-----PKD-----NLNKPYLFLSVLLGAWLGNPGRSQRKVGVAMV 126  
         :||:|||||::|::|::|::|:  
Db     61 GSLIHPQWLTAACHCVGPDKVAALARQLREQHLX-----YQDLPLPVSR I 107  
  
QY    127 EPHP-VYSWEKGCADIALVRLESIQSFSERVLPICLPDSIHILPNPTHWCISWGSGTOD 185  
         ||: |:|||||::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     108 IVHPQFYTAGIG--ADIALLEPEVPKSSHVHTVTLPPASETTPPGPCPWTVTGWDVDN 165  
  
QY    186 GVPLPHQPOTLKLVPIIDSEVCSHLYWRGAGQG----ITEDMLCAGYLEGEDRACLGD 241  
         ||: |:|||||::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     166 DERLPPFPLEKQVKVPMENHIHCDAKHVLGAYTGDDRVIRVDDMLCAG--NTRDRSCOGD 223  
  
QY    242 SGGPLMCQDVGWALLAGIIWSGECAERNRPVVISISAHSRWVEKIV 289  
         :|||::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     224 SGGLVCKVNTWLQAGVSWGEGCAQPNRPGVIETRYTYLDWIHHVY 271

RESULT 15

TEST\_MOUSE STANDARD; PRT; 324 AA.

ID TEST\_MOUSE ID AC Q8JHJ7: O9DAL4:

DT DT 16-OCT-2001 (Rel. 40, Created)

DT DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE DE 01-MAR-2002 (Rel. 41, Last annotation update)

TE Testisin precursor (EC 3.4.21.-) (Trypsinase 4).

GS PRSS21.

ON Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=129/SV;  
RX MEDLINE=21153229; PubMed=11231276;  
RA Scarnan A.L., Hooper J.D., BoucAUT K.J., Sit M.-L., Webb G.C.,  
Normyle J.F., Antalıs T.M.;  
RT "Organization and chromosomal localization of the murine Testisin gene encoding a serine protease temporally expressed during spermatogenesis.";  
RL Eur. J. Biochem. 268:1250-1258(2001).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BALB/c; TISSUE=Testis;  
RC PubMed=11259427;  
RX Wong G.W., Li L., Madhusudhan M.S., Krillis S.A., Gurish M.F.,  
RO Rothenberg M.E., Salı A., Stevens R.L.;  
RA "Trypsinase 4, a new member of the chromosome 17 family of mouse serine proteases";  
RL J. Biol. Chem. 276:20648-20658(2001).  
[3]  
RN SEQUENCE OF 3-324 FROM N.A.  
RP STRAIN=C57BL/6J; TISSUE=testis;  
RC MEDLINE=21085660; PubMed=11217851;  
RX Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okasaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RADOTA K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W.G., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo F., Nikaido I., Pesole G., Quackenbush J.,  
RASchriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojnuna N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RALyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA [3]

Query Match 30.8%; Score 530.5; DB 1; Length 275;







---

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 10:21:17 ; Search time 30.16 Seconds  
(without alignments)  
1818.282 Million cell updates/sec

Title: US-09-903-925A-263  
Perfect score: 1720  
Sequence: 1 MWVGAPALGGGLGFTS.....AQGGALRAPSGSGAARS 317

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP TREMBL.19.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phage.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_rviro.\*
- 17: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1213.5	70.6	297	11	088781
2	660.5	38.4	389	13	Q9PVX7
3	603	35.1	317	13	Q9DGR3
4	578.5	33.6	339	11	Q99L44
5	571	33.2	310	11	Q91XC4
6	570	33.1	321	4	Q96R28
7	564	32.8	310	11	Q9QY29
8	559.5	32.5	273	6	Q9XSM1
9	537	31.2	799	11	Q9PB10
10	536	31.2	237	6	Q29464
11	529.5	30.8	329	13	O42272
12	524	30.5	273	11	Q921N4
13	517.5	30.1	275	4	Q96R26
14	491	28.5	806	6	O18783
15	485.5	28.2	279	11	Q99MS4
16	480	27.9	282	11	Q9D4I3

17	480	27.9	322	11	Q920S2
18	475	27.6	624	11	Q91Y47
19	473	27.5	624	11	Q9DAT3
20	472.5	27.5	405	4	Q96E86
21	469	27.3	624	6	Q95ME7
22	463.5	26.9	267	5	Q9BK47
23	462.5	26.9	643	6	Q97506
24	462	26.9	492	4	Q96T73
25	455.5	26.5	812	11	Q9R0W3
26	452	26.3	454	6	O46506
27	450	26.2	334	6	O46507
28	445.5	25.9	812	11	Q91WJ5
29	445.5	25.9	1524	13	Q91674
30	445	25.9	787	5	Q9VEY6
31	445	25.9	810	4	Q15146
32	439.5	25.6	490	11	Q920K3
33	437.5	25.4	367	11	O70169
34	435.5	25.3	581	4	Q9BYE2
35	435	25.3	263	11	Q9DC86
36	435	25.3	277	5	O96899
37	434	25.2	265	5	Q17800
38	433	25.2	329	6	Q9GL10
39	433	25.2	537	4	Q9BYE1
40	432.5	25.1	767	13	Q9DGR2
41	430.5	25.0	266	11	O70170
42	430	25.0	263	11	Q9CR35
43	429.5	25.0	261	13	Q9W7Q4
44	429.5	25.0	279	11	Q9QZ74
45	428.5	24.9	249	13	Q9W6K0

## ALIGNMENTS

RESULT 1  
088781  
ID 088781 PRELIMINARY; PRT; 297 AA.  
AC 088781;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SERINE PROTEASE PRECURSOR (FRAGMENT).  
GN BSP2.  
OS Rattus rattus (Black rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=BRAIN;  
RX MEDLINE=98389725; PubMed=9722524;  
RA Davies B.J., Pickard B.S., Steel M., Morris R.G., Lathe R.;  
RT "Serine Proteases in Rodent Hippocampus."  
RL J. Biol. Chem. 273:23004-23011(1998).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
CC ENBL; AJ005642; CAA06644.1; -.  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.252; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF000089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease; Signal.  
FT NON\_TER 1  
FT SIGNAL <1 23 POTENTIAL.  
FT CHAIN 24 297 SERINE PROTEASE.  
SQ SEQUENCE 297 AA; 32086 MW; 6DA25C6633D6AB55 CRC64;

Q920s2	mus musculus
Q91y47	mus musculus
Q9dat3	mus musculus
Q96e86	homo sapien
Q95me7	oryctolagus
Q9bk47	luidia foli
Q97506	sus scrofa
Q96t73	homo sapien
Q9r0w3	rattus norv
O46506	papio hamad
O46507	papio hamad
Q91wj5	mus musculus
Q91674	xenopus lae
Q9vey6	drosophila
Q15146	homo sapien
Q920k3	rattus norv
O70169	mus musculus
Q9bye2	homo sapien
Q9dc86	mus musculus
O96899	scolopendra
Q17800	caenorhabdi
Q9gl10	ovis aries
Q9bye1	homo sapien
Q9dgr2	xenopus lae
O70170	mus musculus
Q9cr35	mus musculus
Q9w7q4	paralichthy
Q9qz74	rattus norv
Q9w6k0	notothenia

Query Match	70.6%	Score 1213.5	DB 11	Length 297
Best Local Similarity	74.7%	Pred. No. 3e-104		
Matches 222	Conservative 26	Mismatches 40	Indels 9	Gaps 1
QY 6	APPALGGCGCTFTT	SEILLASTAILNARIPVPPACGKPOQLNRVVGEDSTDSWPWIV 65		
DB 5	SPGL-----TLFILLPSATVSAANIRGSPDCGKPOQLNRVVGEDSADQWPWIV 55			
QY 66	STQKNGTHHCAGSLT	SRWVITAAHCFKDLNKPVLFVLLGAMOLGNPGRSOKVGVAV 125		
DB 56	STLKNSSHHCAGSLT	NRVVVSAHCFSSNDKSPSYVLLGAKLGNPGRSOKVGVAS 115		
QY 126	VEPHVYVSWKGCACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGWSIQD 185			
DB 116	VLPHPVYRKEGTHADIALVRLERPIQFSERILPICLPDSSVHLPPNTNCWIAWGSIQD 175			
QY 186	GVPLPHTLOKLVPIIDSEVCSHLWRGAGQPIEDMLCAGLYGERDACLDGSGGP 245			
DB 176	GVPLPHTLOKLVPIIDPECKSLYWRGAGQEAITEDMLCAGLYGERDACLDGSGGP 235			
QY 246	LMCQVDGAWLLAGIISWEGCAERNRPGVYISLSAHRSWVEKIVQGVQLRGAOGG 302			
DB 236	LMCQVDHLLTGIISWEGCAERNRPGVYISLSAHRPWVQRIQGVQLRGLADSG 292			
RESULT 2				
Q9PVX7				
ID	Q9PVX7	PRELIMINARY;	PRT;	389 AA.
AC	Q9PVX7;			
DT	01-MAY-2000 (TremBrel. 13, Created)			
DT	01-MAY-2000 (TremBrel. 13, Last sequence update)			
DT	01-DEC-2001 (TremBrel. 19, Last annotation update)			
CC	EPIDERMIS SPECIFIC SERINE PROTEASE.			
GN	XEPSIN.			
OS	xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCHI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RT	Yamada K.;			
RT	"The expression control of xepsin by non-axial and planar			
RT	posteriorizing signals in Xenopus epidermis.";			
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
DR	EMBL; AB018694; BAA84941.1; -.			
DR	HSP; P00763; IDPO.			
DR	MEROPS; S01 UPA; -.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Trypsin.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; Tryp_Spc; 1.			
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease.			
SQ	SEQUENCE 389 AA; 42375 MW; B31PB4A2F5D1F6E3 CRC64;			
Query Match	38.4%	Score 660.5	DB 13	Length 389;
Best Local Similarity	47.0%	Pred. No. 6.2e-53		
Matches 131	Conservative 42	Mismatches 95	Indels 11	Gaps 6
QY 40	ACGKPOQLNRVVGEDSTDSWPWIVSIQKNGTHHCAGSLT	SRWVITAAHCFKDLNKP 99		
DB 16	ACGVPVLSNRVVGMSKRGFWPQISLSYKSDICGSLTDSWVWTAHCHI-DSLDVS 74			
QY 100	YLFVLLGAMOLGNPGRSOKVGVAVVEPHVYVSWKGCACADIALVRLERSIOFSRVL 159			



```
Query Match      33.1%; Score 570; DB 4; Length 321;
Best Local Similarity 41.1%; Pred. No. 1.2e-44;
Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

QY 15 LGTTSLLASTAILNAARIPVPACGKPOQLN---RVVGEDSTSEWPIVSIQKNG 71
DQ 15 LGTTSLLASTAILNAARIPVPACGKPOQLN---RVVGEDSTSEWPIVSIQKNG 71
DQ 3 LGAGCLLLLL---AVPGVSLRTLQGGCRPOVSDAGGRIIVGGHAAPAGAWPQASLRLR 59
QY 72 THHCAGSLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHV 131
DQ 72 THHCAGSLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHV 131
DQ 60 VHVCGSLLSPQWLTAHCFSGSLNSD-YQVHLGELETILSPHFT---VROILHSS 115
QY 132 YSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPH 191
DQ 132 YSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPH 191
DQ 116 PSGQPGTSGDIALVELSPVTLSSRIPLVCLPEASDDFCGICRCWVGWGTREGEPLP 175
QY 192 POTLQKLVPIIDSEVCSHLYWRGAGQPTEDMLCAGYLEGERDACLGDGSGGLPQCOVD 251
DQ 192 POTLQKLVPIIDSEVCSHLYWRGAGQPTEDMLCAGYLEGERDACLGDGSGGLPQCOVD 251
DQ 176 PYSLVREKVSVDYETCRDQY-PGPGGSILOQDMLCA---RGPDAQDDSGGLPVCQVN 231
QY 252 GAWLLAGTISWEGCAERNRPGVYISLSAHSRWVEKIVQGVQLRGAQGG---GALRAP 307
DQ 252 GAWLLAGTISWEGCAERNRPGVYISLSAHSRWVEKIVQGVQLRGAQGG---GALRAP 307
DQ 232 GAWVQAGISWEGCGRPNRPGVYTRPVAVVNIIRRI-----TASGGSSEGYPLRP 283

RESULT 7
QYQY29 PRELIMINARY; PRT; 310 AA.
AC QYQY29 PRELIMINARY; PRT; 310 AA.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DISTAL INTESTINAL SERINE PROTEASE.
GN DISP OR DISP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20246299; PubMed=10786627;
RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,
RA Walters J.R.;
RT "Characterization of a novel murine intestinal serine protease,
RT DISP.";
RL Biochim. Biophys. Acta 1490:131-136(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ243866; CAB56465.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.042; -.
DR MGD; MGI:1353645; Disp.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp.SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
FT CHAIN 29 273 TRYPTASE.
SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;

Query Match      32.8%; Score 564; DB 11; Length 310;
Best Local Similarity 42.4%; Pred. No. 4e-44;
Matches 117; Conservative 48; Mismatches 95; Indels 16; Gaps 9;

QY 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---IQKNGTHCAG 77
DQ 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---IQKNGTHCAG 77
DQ 2 LHLALALLSL--VSNAPGQALQSGIIGKEAPGSRWPQVSLVRDQYWRHCCG 59
QY 78 SLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHVSIQKNG 137
DQ 78 SLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHVSIQKNG 137
DQ 60 SLTHPQWLTAHCFKDPQLQEPDSDFRVLREQHL---YYQDRLLPISRVPHPHYVVEN 116
QY 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPHQTLQK 197
DQ 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPHQTLQK 197
DQ 10 LLLL---QILTRAGDILPVCVCHSRDAGKIVGGQDALEQWFWQVSLWITEDG-HICGG 65

Query Match      32.5%; Score 559.5; DB 6; Length 273;
Best Local Similarity 40.9%; Pred. No. 8.8e-44;
Matches 113; Conservative 53; Mismatches 95; Indels 15; Gaps 7;

QY 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---KNGTHCAG 77
DQ 21 LLLASTAILNAARIPVPACGKPOQLNVRVVGEDSTSEWPIVSIQ---KNGTHCAG 77
DQ 2 LHLALALLSL--VSNAPGQALQSGIIGKEAPGSRWPQVSLVRDQYWRHCCG 59
QY 78 SLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHVSIQKNG 137
DQ 78 SLTTSRWITAAHCFKDNLNKPYLFVLLGAWQLNPGSRQKVGAVVPHVSIQKNG 137
DQ 60 SLTHPQWLTAHCFKDPQLQEPDSDFRVLREQHL---YYQDRLLPISRVPHPHYVVEN 116
QY 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPHQTLQK 197
DQ 138 ACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGWGSIQDGVPLPHQTLQK 197
DQ 117 G-ADIALQLLEEPSVTSCHVRPVTLPASFTFPFPPSQCWWTGNGVNDNGRPLPPVPLKQ 175
```

[illegible]

[illegible]

Df 177 NVKILSRQTCNCLYHINPSSDLSGVSVOODMICAGSAAAGSVDACQGDGGGLTCTVNNQPY 233

Qy 256 LAGIISWEGCGCAERNRPBGVIYSLSAHSRWVEKI 288  
II : :: III : I : IIIIII : : II : I :

Df 237 LAUVSGDECGAPNRPGVILISLYSWIRSI 269

RESULT 12

ID Q921N4 PRELIMINARY; PRT; 273 AA.

AC O921N4;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MAST CELL PROTEASE 7.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;

RX [1]

RN SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; BC011328; AHI1328.1; -.

KW Protease.

SQ SEQUENCE 273 AA; 30332 MW; 51417476514035BE CRC64;

Query Match . 30.5%; Score 524; DB 11; Length 273;  
Best Local Similarity 38.1%; Pred. No. 1.7e-40;  
Matches 109; Conservative 55; Mismatches 84; Indels 38; Gaps

Qy 22 LLLASTAILNAARIPVPACGKPQLNRVVGGSDTSSEWPMTIVSIQNKT---HHCCGS 78  
||| : : :: : | | : ||| : : : ||| : : : | | | |

Df 4 LLLLTPLSLSLVAAP--GPAMTRFGIVGGGEAHCKNPQVSLRANDTYMHFCGSGS 60  
|||| : : : : : | | : ||| : : : ||| : : : | | | |

Qy 79 LLTSRWVTIAHC-----FKONLNKPYLFSLVLGAWLGNPCRSQKVGVAVWPEP 128  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Df 61 LIHPDWLTAACHCVPDPADPNKVRQLKOYL-----YHDHLMTVSQIIT 107

Qy 129 HP-VYSWKECACADIALVRLERSIQSERVLPICLPDASIHLPPTHCHWSIGWSTODGY 187  
|| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Df 108 HPDPIYVDQG--ADIALLKLTVTNVNISDYVHYPLPPAPSETFSPGTLCWWTGNGINDGV 165  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Qy 188 PLRPHPTQLOLKVIPDIISWCESHLNYRGAGQP----TFEDMLCACYLEGERDACLGDSG 243  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Df 166 LLPFPFLKEVQVPIENHLCDDIKYHKGLTIGNHVIRDMDLCAEN-EG-HDCSCGDGSG 233  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Qy 244 GPLMCQVDGAWLLLAGIISWEGCGCAERNRPBGVIYSLSAHSRWVEKI V 289  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

Df 224 GPLVKVEDTWLQGVVSWGEGCAQPNRPGIYTRYVYILDWIHRVY 269  
||| : : ||| : : : | | : ||| : : : ||| : : : | | | |

RESULT 13

ID Q96RZ6 PRELIMINARY; PRT; 275 AA.

AC Q96RZ6;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE TRYPTASE I.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
NCBI\_TaxID=9606;

RX [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=21056910; PubMed=11157797;

RA Daniels R.J., Feden J.F., Lloyd C., Horsley S.W., Clark K.,  
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,  
RA Higgs D.R.;

RT "Sequence, structure and pathology of the fully annotated terminal 2





```

Db 62 CGGSIHPQWLTAAHCIREDDADPSVPRIRVGEAYLYGG----KELLSVSRVIIHPDFV 117
Qy 134 WKEGACADIALVRLERSIQFSERVLPICLPDASIHLPENTHCWISGWGSIQDGVPLPHQ 193
Db 118 -HAGIGSDVALLQLAVSVQSPFNKPKVLPSESLVTKKDVCTWGTGAVSTHRSPLPPY 176
Qy 194 TLQKLKVPIDSEVCSHLY-----WRGAGQGPTEDMLCAGYLEGERDACLGDSGGPLMC 248
Db 177 RLQOVQVKIIDNSLCEEMYHNATRHNRGQKLIKMLCAGN-QGQ-DSCYGDSSGGPLVC 234
Qy 249 QVDGAWLLAGTISWEGCAERNRPGVYISLSAHSRWEKIVQ 290
Db 235 NVTGSWTLGVVSWGYGCALRDFEGVYARVQSFPLPWITQOMQ 276

```

: Search completed: August 29, 2002, 10:25:30  
: Job time: 253 sec